

22. Anatomía y fisiología del habla

Los órganos articulatorios

Los órganos articulatorios son las partes del cuerpo que intervienen en la producción del habla. La denominación «órganos articulatorios» no es del todo apropiada, ya que su cometido principal consiste en cumplir con las necesidades biológicas básicas de respirar y comer. Pero parece que ha habido un considerable desarrollo evolutivo de su forma, que les permite funcionar eficazmente en la producción del habla (§49).

Probablemente el rasgo más asombroso de cualquier diagrama de los órganos articulatorios sea la cantidad de partes del cuerpo que intervienen en el habla. No se trata simplemente de la boca y la garganta; también tenemos que mostrar cómo se ven implicados los *pulmones*, la *tráquea* y la *nariz*. Dentro de la boca hay que distinguir la *lengua* y las distintas partes del *paladar*. Dentro de la garganta tenemos que distinguir la parte superior, o *faringe*, de la inferior o *laringe*, que contiene los *repliegues vocales*, corrientemente conocidos como «cuerdas vocales». La faringe, la boca y la nariz forman un sistema de áreas huecas, o *cavidades*, conocido como *tracto vocal* (si bien esta denominación a veces incluye también a la laringe y a los pulmones). Cuando movemos los órganos del tracto vocal, cambiamos su forma, y esto es lo que nos permite producir el gran número de sonidos diferentes que tiene el lenguaje hablado.

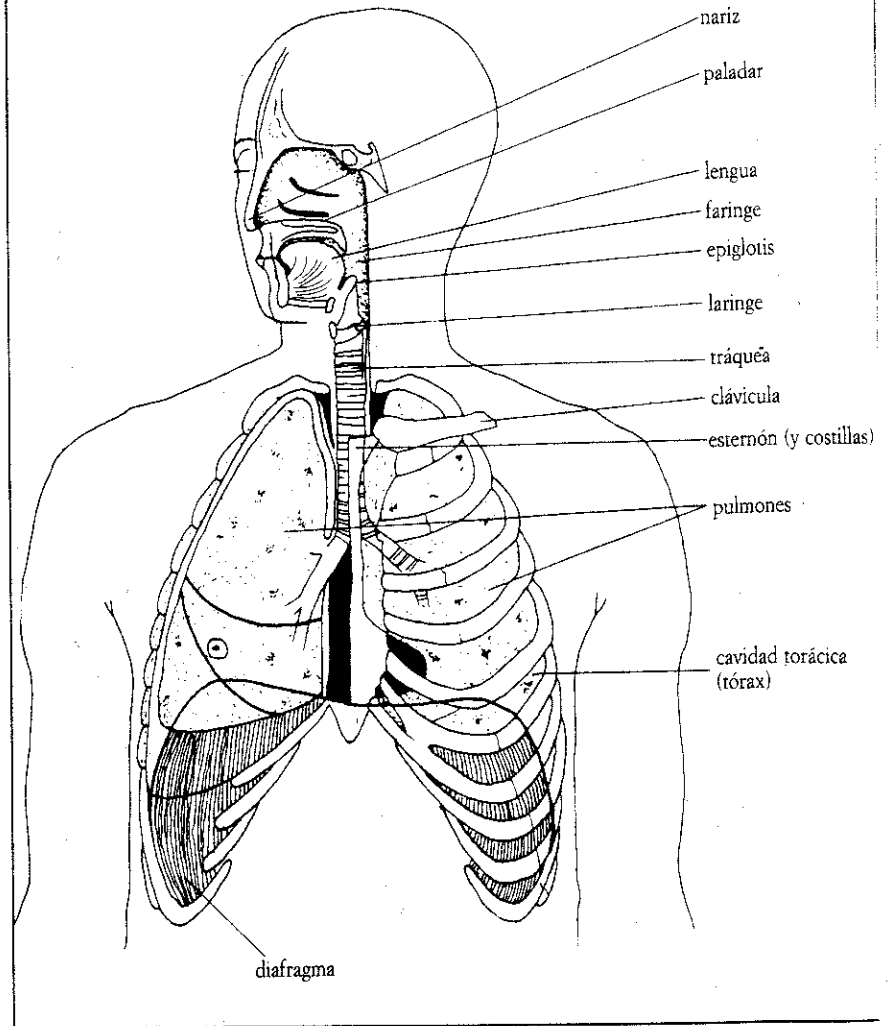
LOS PULMONES

Una condición previa a la producción de cualquier sonido es la existencia de una fuente de energía (pág. 132). En el habla, la energía toma la forma de una corriente de aire que, en circunstancias normales, ha sido puesta en movimiento por los pulmones.

Los pulmones se encuentran en una cavidad del pecho (o *tórax*), conocida como *cavidad torácica*. Dicha cavidad está delimitada en su parte posterior por la columna vertebral, en la anterior por las costillas y el esternón, y en la inferior por el músculo con forma de bóveda conocido como *diafragma*, que separa los pulmones de las cavidades inferiores del abdomen. La estructura que rodea la cavidad torácica se conoce como *caja torácica*. El acto de la respiración se produce gracias a la acción de la caja torácica, que permite a los pulmones funcionar como fuelles, dejando que el aire fluya hacia dentro y hacia fuera.

Para hablar, primero tenemos que inhalar. Una señal que parte del nervio central en el tronco cerebral (desde donde se controla la respiración) provoca la contracción de los músculos de la caja torácica: concretamente, los músculos que están entre las costillas (los músculos *intercostales*) hacen que las costillas se muevan hacia arriba y hacia fuera, y que el diafragma se mueva hacia abajo. El resultado es que el pecho se expande, y así también los pulmones, lo que provoca que la presión del aire que es-

Disposición general de los órganos vocales



tos contienen disminuya momentáneamente. Inmediatamente entra aire en los pulmones, para igualar la presión con la de la atmósfera exterior.

Después espiramos. Contraemos el pecho, de modo que los pulmones, al bajar las costillas y subir el diafragma, obligan al aire a salir. Pero nunca expulsamos todo el aire. Cuando mantenemos una conversación normal sólo usamos una cuarta parte del aire de los pulmones, si bien la cantidad aumenta en cierto grado si hablamos más alto y con más esfuerzo, como al gritar, al actuar, al pronunciar discursos en público o al hacer un «aparte» en el teatro.

USAR LOS PULMONES PARA HABLAR

El aire de los pulmones se suele denominar *aire pulmonar*. Cuando el aire pulmonar sale hacia fuera se dice que es *egresivo*. La gran mayoría de los sonidos del habla se producen utilizando aire pulmonar egresivo.

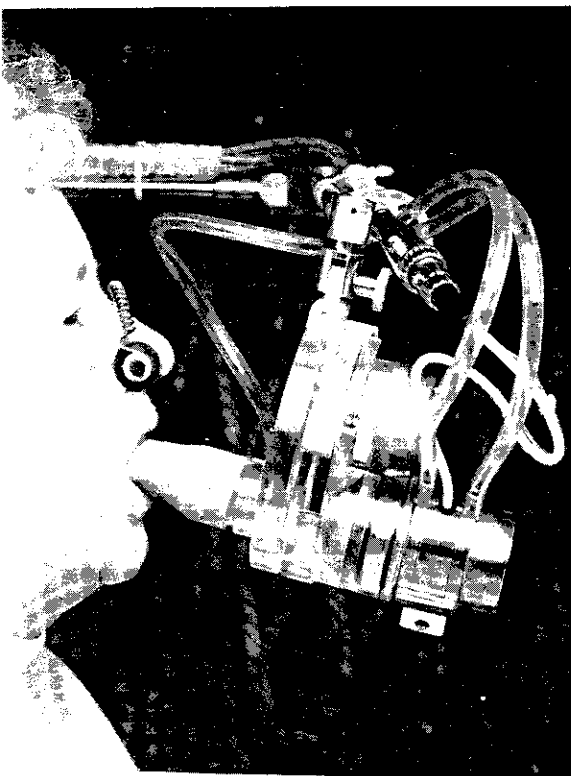
También es posible (aunque no usual) hablar mientras la corriente de aire entra hacia los pulmones (*aire pulmonar ingresivo*). Se usa esta corriente de aire en algunas ocasiones, cuando alguien intenta hablar mientras ríe o llora, o cuando está sin aliento. Algunas palabras breves a veces se pronuncian con una corriente de aire ingresivo, cuando usamos un tono de voz de «rutina» para contestar a lo que alguien está diciendo. Puede percibirse un uso alternado de aire egresivo e ingresivo cuando alguien está contando muy deprisa, en voz baja. Pero el habla «ingresiva» es de pobre calidad, suena como un graznido apagado, y para mucha gente resulta desagradable. No entra en el uso corriente del español de cada día.

El ciclo respiratorio

Los acontecimientos que se suceden cuando inhalamos (*inspiración*) y exhalamos (*expiración*) reciben el nombre

Neumotacógrafo (derecha). Dentro de la mascarilla de este aparato hay distintos contadores que señalan el volumen de aire que sale por la boca y por la nariz.

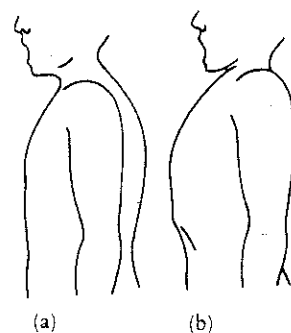
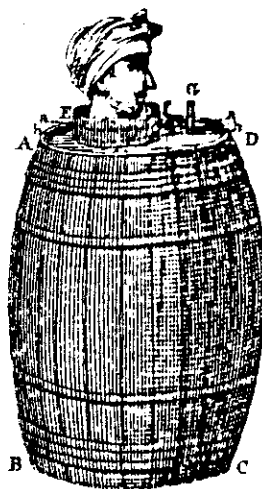
Espirómetro (abajo). El espirómetro mide el volumen de aire generado por los pulmones. La información queda grabada en un cilindro giratorio.



de *ciclo respiratorio*. Normalmente, las dos mitades de este ciclo tienen aproximadamente la misma duración; pero cuando estamos hablando, se pasa a una sucesión de inspiraciones rápidas y espiraciones lentas. También cambia el ritmo al que respiramos. Cuando estamos callados y en reposo, respiramos una media de 12 veces por minuto, de modo que el tiempo que tardamos en inspirar y espirar es aproximadamente de dos segundos y medio para cada fase. Al hablar, reducimos el tiempo de inspiración a tan sólo un cuarto de segundo, y generalmente extendemos el tiempo de espiración a 5 o 10 segundos, incluso a 20 segundos a veces, según cuál sea el control sobre la propia voz, el estado emocional, y otros factores semejantes. Esta alteración de nuestro ritmo respiratorio permite que nuestras espiraciones «lleven» una cantidad de habla mucho mayor de lo que sería posible en el caso contrario. En una conversación corriente es completamente normal articular entre 250 y 300 sílabas en un minuto.



Un pletismógrafo del siglo XVIII. La pletismografía es una técnica que sirve para determinar el tamaño de una parte del cuerpo, y se puede usar para estudiar el volumen de los pulmones. Este pletismógrafo consistía sencillamente en un barril con un collarín hermético para el cuello. El acto de respirar altera el volumen del cuerpo del sujeto, lo que a su vez altera la cantidad de aire dentro del barril. Un pequeño tubo en la parte de arriba permite el paso del aire. El seguimiento de los cambios en el paso del aire a través del tubo proporciona una medida del cambio de volumen de los pulmones. Hoy en día todavía se utiliza una forma desarrollada de este mismo método.



Formas de respirar correctas e incorrectas

Las técnicas de control de la respiración constituyen una parte fundamental de cualquier programa de formación para mejorar el uso de la voz, como en el caso del canto, el teatro o las terapias del habla (§46). El tipo de respiración más eficaz requiere una inspiración rápida y una espiración controlada, que corresponda a las necesidades de la voz. Un cuidadoso control del movimiento de las costillas y del diafragma es el rasgo principal del método de respiración *intercostal diafragmática*, considerado por los especialistas como el más eficaz en la mayoría de los casos. Contrasta con varios métodos menos eficaces, como la forma de respiración tensa y nerviosa, en la parte superior del tórax, que se conoce como respiración *clavicular*. Si se la utiliza habitualmente en el habla, la excesiva tensión muscular que se necesita para mantener este tipo de respiración puede dañar las cuerdas vocales, causando ronquera y otras variedades de voz anormales. Las ilustraciones de arriba muestran las posiciones del cuerpo correcta (a) e incorrecta (b) para cantar.

Hablar sin los pulmones

Todas las vocales y consonantes del español, al igual que las de la mayoría de las lenguas, se articulan utilizando aire pulmonar egresivo. Pero existen varios otros tipos de sonidos del habla que no usan aire de los pulmones, y que están presentes en muchas lenguas del mundo.

CLICS

Uno de los tipos más característicos de sonido no pulmonar es el clic. Los sonidos clic son ruidos secos de succión, que se hacen con la lengua o los labios. Por ejemplo el sonido que a menudo aparece representado como *tsk tsk* en los comics está formado por un par de sonidos clic que se producen chasqueando la lengua contra los dientes superiores. Al producir un sonido clic se puede respirar hacia fuera o hacia dentro, independientemente, lo que demuestra que los pulmones no participan en el proceso.

En las lenguas europeas, se pueden oír con frecuencia sonidos clic aislados, que equivalen a sonidos con significado, pero no forman parte del sistema de vocales y consonantes (§28). El clic mencionado anteriormente, por ejemplo, expresa desaprobación en español, pero es un sonido que no se utiliza como parte de una palabra, igual que /t/ o /p/. No obstante, en muchas otras lenguas los clics sirven como consonantes. Las más conocidas son algunas de las lenguas del sur de África, que se suelen denominar «lenguas clic». El cunگو es una de esas lenguas, con más de 48 clics (pág. 168). Las lenguas joisanas, que incluyen las lenguas de las tribus joiói (hotentotes) y san (bosquimanos), son las que tienen un sistema de clics más complejo, ya que utilizan muchos puntos de articulación diferentes en la boca, y añaden otros sonidos producidos en la garganta o la nariz.

SONIDOS GLOTALES

El espacio que hay detrás de la nuez, entre las cuerdas vocales, se llama *glotis*. Podemos usar la glotis para poner en movimiento una corriente de aire, y varias lenguas utilizan sonidos basados en este principio, que se conoce como mecanismo de aire *glotal*. Cuando la glotis provoca un movimiento de aire hacia dentro, los sonidos se llaman *inyectivos*. Una consonante inyectiva es un sonido glotal ingresivo. Cuando el aire se mueve hacia fuera, los sonidos se llaman *eyectivos*. Una consonante eyectiva es un sonido glotal egresivo.

Existen consonantes inyectivas en muchas lenguas, pero son especialmente frecuentes en las lenguas amerindias y africanas (como el chona y el iyo). Las consonantes eyectivas se utilizan ampliamente en las lenguas de la familia caucásica, y también en muchas lenguas amerindias y africanas (como el hausa y el amárico). Se las puede oír incluso en ciertas variedades del inglés. Los hablantes del norte de Inglaterra las usan a menudo como final de palabra, en lugar de las normales /p/, /t/, /k/ pulmonares.



Miriam Makeba. Sus grabaciones de «canciones clic» se hicieron muy populares en la década de 1960. Hablante de josa como lengua materna, utilizaba en sus canciones palabras que contenían consonantes clic, consiguiendo un efecto notable al articularlas con gran resonancia.

Cómo se producen los sonidos clic

La producción de un sonido clic tiene lugar sólo en la boca. El paso del aire se controla mediante movimientos efectuados contra la parte posterior del cielo de la boca, es decir, el *velo* del paladar (pág. 130). Por esta razón se dice que los sonidos clic utilizan un mecanismo de aire *velar*. Lo que a veces se representa como *tsk tsk* es un ejemplo de clic *dental* doble (símbolo fonético [ɽ]), porque los dientes participan en su articulación. En todo el oriente próximo, un solo clic

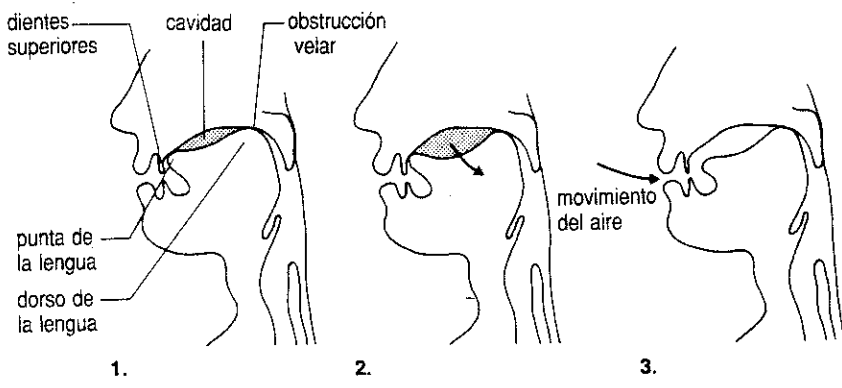
dental se usa para expresar negación. Los clics laterales (símbolo fonético [ɬ]) se producen con los bordes de la lengua, y corresponden a los chasquidos con los que arreamos a los caballos y a otros animales (incluyendo a los humanos). Un clic hecho con los labios se llamaría clic *bilabial*; hecho con los labios fruncidos, se usa a menudo como «beso a distancia».

1. El postdorso de la lengua está levantado, de manera que presiona contra el velo. Al mismo tiempo, en

la parte anterior de la boca se produce una oclusión, mediante la lengua o los labios. Esto forma en la boca una cavidad, aislada del exterior.

2. El dorso de la lengua se mueve ligeramente hacia abajo y hacia atrás, como para formar un vacío parcial dentro de la cavidad.

3. Cuando se baja la lengua o se abren los labios de forma repentina, rápidamente entra aire desde fuera, produciendo el sonido que oímos como un clic.



Cómo se producen los sonidos eyectivos

El rasgo fundamental de los sonidos eyectivos es que la glotis permanece firmemente cerrada, de manera que no puede pasar aire desde o hacia los pulmones. De hecho, lo que hacemos es «aguantar la respiración» durante un breve instante.

lengua. De esta manera una cierta cantidad de aire queda atrapada entre la glotis y la obstrucción de la parte superior del tracto vocal.

la lengua, y el sonido «escapa» hacia fuera.

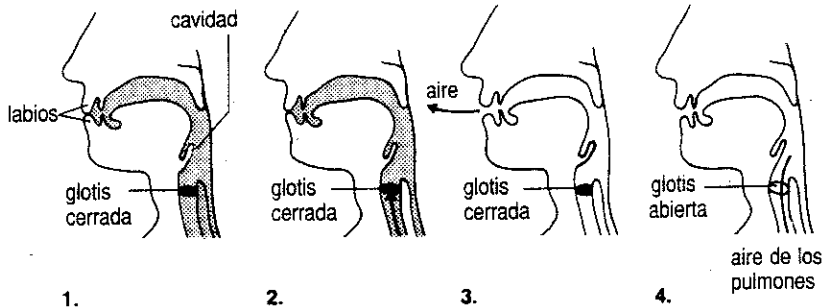
4. La glotis se abre y por el tracto vocal sube rápidamente aire de los pulmones, que actuará como fuente de energía para el próximo sonido del habla.

Todo el proceso, desde la obstrucción inicial de la glotis hasta su apertura final, dura, como media, sólo una veintésima de segundo, aunque el tiempo varía mucho según las lenguas.

1. Al mismo tiempo que cerramos la glotis, nos preparamos para articular un sonido consonántico: para la [p] cerramos los labios, para la [t] o la [k] elevamos la

2. Contraemos algunos de los músculos de la laringe, para hacer que la glotis se mueva hacia arriba, movimiento que comprime el aire de la cavidad.

3. La creciente presión se libera de golpe al eliminar la obstrucción en la boca, abriendo los labios o bajando



Cómo se producen los sonidos inyectivos

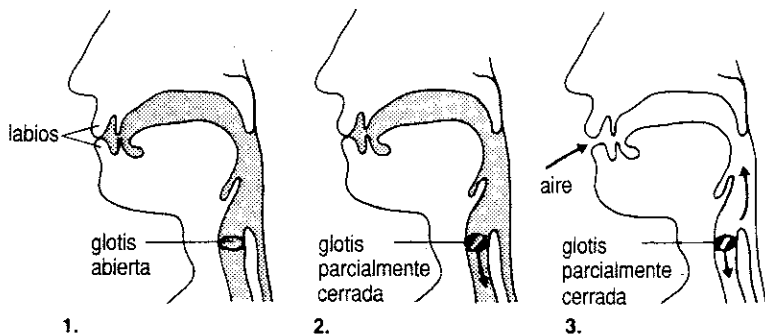
El proceso que se emplea para los sonidos inyectivos es en gran medida el opuesto del que se emplea para los eyectivos.

2. Los músculos de la laringe se usan para cerrar parcialmente la glotis y moverla hacia abajo, de manera que la presión de la cavidad por encima de la glotis se ve algo reducida. La glotis no queda completamente cerrada (al contrario que en los sonidos eyectivos), de manera que todavía puede pasar entre las cuerdas vocales una cierta

cantidad de aire, que las hace vibrar.

3. Cuando abrimos los labios o hacemos descender la lengua, desaparece la oclusión del tracto vocal, y la boca absorbe aire del exterior. Este se mezcla en la glotis con el aire de los pulmones, para producir un sonido con una resonancia sorda y apagada.

1. Provocamos una obstrucción en el tracto vocal, para la [b] con los labios, para la [d] o la [g] con la lengua. Nótese los símbolos fonéticos especiales, para distinguir estos sonidos de las [b], [d] y [g] pulmonares.



El tracto vocal puede producir muchas más clases de sonidos, pero estos no se usan con regularidad en el lenguaje hablado. Sonidos como los que se producen si hacemos rechinar los dientes, si hacemos chasquear la lengua contra el fondo de la boca, o si producimos un ruido de aspiración colocando la lengua contra el interior de la mejilla, se consideran peculiaridades del hablante. El oyente no suele interpretarlos como intentos de comunicación.

Por otra parte, cuando las personas se comunican, usan también otros mecanismos para generar una corriente de aire. Un sonido velar egresivo (con el mismo mecanismo que el de un clic, pero expulsando el sonido hacia fuera, y no succionándolo hacia dentro), producido con los labios, es bastante común en francés, donde, acompañado de un gesto distintivo de la mano y un encogimiento de hombros, significa aproximadamente «Y a mí qué» o «No es culpa mía». Un sonido semejante, pero con la lengua sobresaliendo un poco, es en muchas lenguas una señal de desprecio: se trata de lo que normalmente llamaríamos «pedorreta».

Los mecanismos anormales de generación de aire también se utilizan en circunstancias especiales. Se puede comprimir aire dentro del espacio de las mejillas y utilizarlo para hablar, obteniendo así la voz que hizo famosa al Pato Donald de Walt Disney. También se pueden generar sonidos empleando aire que proviene del estómago o del esófago (conducto que va desde la faringe al estómago), como en un eructo. La voz *esofágica* la emplean, de forma artificial, las personas a las que les ha sido extirpada la laringe por una enfermedad (pág. 276).

Es normal que una lengua utilice sólo uno o dos mecanismos de producción de aire para la articulación de vocales y consonantes. Todas las lenguas utilizan aire pulmonar egresivo. También está muy extendido (aunque no en las lenguas europeas) el uso de aire glotal egresivo (para los sonidos eyectivos). Mucho menos frecuente es el aire glotal ingresivo (para los inyectivos), y sólo un pequeño grupo de lenguas africanas emplea el aire velar (para los clics). Es poco corriente encontrar una lengua que use regularmente más de uno o dos de estos mecanismos; sólo unas pocas utilizan tres. Un caso único es el del damín, una lengua ritual de una tribu aborigen del norte de Australia, los lardil, de la que se dice que utiliza por lo menos cinco mecanismos distintos para producir corrientes de aire. Esta lengua usa sonidos pulmonares egresivos, glotales egresivos y velares ingresivos, pero además tiene un sonido [l] pulmonar *ingresivo* y un sonido [p] velar *egresivo*. No se ha descubierto ninguna otra lengua cuyas consonantes impliquen este último tipo de sonidos, lo que ha llevado a algunos especialistas a suponer que tal vez el sistema de sonidos de esta lengua haya sido inventado especialmente para cumplir una función ritual.

Antes de que podamos hablar, el aire de los pulmones debe transformarse en vibraciones audibles, mediante la utilización de los distintos órganos del tracto vocal. La fuente de vibraciones más importante para la producción de los sonidos del habla es la parte inferior del tracto, la laringe.

La laringe está situada en la parte superior de la tráquea. Consiste en un tubo formado por cartílagos conectados por membranas y ligamentos, dentro del cual se encuentran las dos bandas de tejido muscular que se conocen como cuerdas vocales. La situación de la laringe puede apreciarse fácilmente, porque la parte anterior, el cartílago *tiroides* (literalmente 'con forma de escudo'), forma en el cuello un ángulo prominente, conocido como la «nuez» o «bocado de Adán» (que sobresale más en el caso de los varones). Otros dos cartílagos contribuyen a definir, junto con el tiroides, el área de la laringe: el cartílago *cricoides* ('con forma de anillo'), y los dos cartílagos *aritenoides* ('con forma de cazo'). Los tres, al moverse, ayudan a controlar el modo en que vibran las cuerdas vocales. A la derecha ilustramos su disposición anatómica.

La apertura entre las cuerdas vocales, conocida como *glotis*, es un área bastante pequeña. En los hombres, el borde interior de las cuerdas mide entre 17 y 24 mm., mientras que en las mujeres es incluso menor, entre 13 y 17 mm. aproximadamente. A los pliegues de la glotis se les suele llamar las *verdaderas* cuerdas vocales, puesto que un poco por encima de ellos existe un segundo estrechamiento de la laringe, conocido como «falsas» cuerdas vocales, o cuerdas «ventriculares». No es normal que se utilicen para la fonación, aunque con frecuencia intervienen en ciertos tipos de timbre de voz (siendo un ejemplo notable la voz áspera del músico de jazz Louis Armstrong), y en el efecto de voz «bi-tonal» que se escucha en ciertos casos de trastornos vocales (pág. 276).

Las cuerdas vocales

Las cuerdas vocales son extraordinariamente versátiles. Pueden variar su tensión, su elasticidad, su altura, su an-

Una perspectiva evolutiva

Desde el punto de vista biológico, la laringe actúa como una válvula que controla el movimiento del aire desde y hacia los pulmones, y evita que comida, cuerpos extraños u otras sustancias puedan entrar en ellos. Además, al cerrar las cuerdas vocales, se puede aumentar la presión dentro de los pulmones, en la forma requerida para todas las formas de esfuerzo muscular (por ejemplo, levantamiento de un peso, defecación, tos).

En el curso de la evolución, la laringe se ha llegado a adaptar para constituir la mayor fuente de sonido para el habla. No obstante, su posición delante de la baja faringe (que conduce al estómago) presenta una complicación, porque supone el paso de la comida y los líquidos por la entrada a la tráquea, en su camino hacia el estómago. (Esta complicación no existe en otros animales, en los que la laringe está situada más arriba (§ 49).) Para resolver el problema, un cartílago con forma de hoja, llamado *epiglotis*, tapa la entrada a la laringe, como parte del mecanismo de deglución, evitando así que esas sustancias tomen la dirección equivocada.

altura, su longitud o su espesor, gracias a la compleja interacción del gran número de grupos de músculos que controlan el movimiento de la laringe. Dichos movimientos se producen muy rápidamente durante el habla, y son los responsables de varios tipos de efectos auditivos.

Fonación. El efecto más importante es el de la producción de una vibración audible, un sonido zumbante, conocido como *sonoridad* o *fonación*. Todas las vocales y la mayor parte de las consonantes (por ejemplo [m], [b], [g]) utilizan este efecto. De hecho, es posible apreciar la vibración, por ejemplo, colocando el índice y el pulgar a ambos lados de la nuez, y comparando el efecto al pronunciar [bbb] y [fff] en voz alta. Otra posibilidad de comprobar el efecto de resonancia de las cuerdas vocales es la de articular esos sonidos mientras nos tapamos las orejas con los dedos.

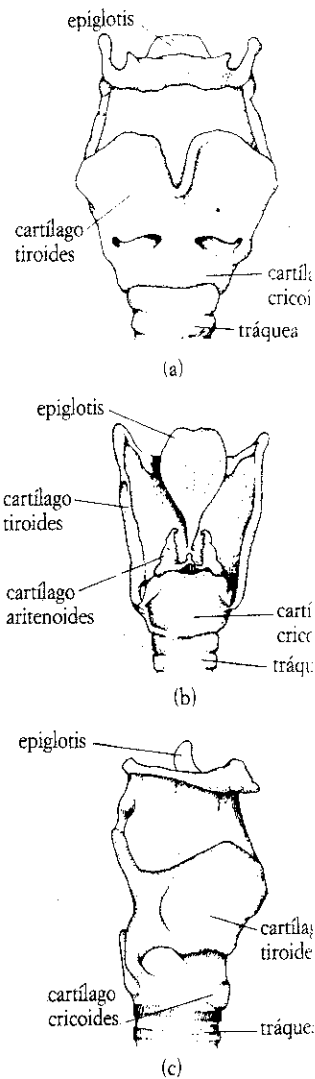
Cada pulsación de la vibración representa un solo movimiento de apertura y cierre de las cuerdas vocales. En las voces masculinas adultas, esta acción se repite una media de aproximadamente 120 veces (o «ciclos») por segundo, que corresponde a una nota en el piano de una octava por debajo del Do central. En las mujeres, la media es algo menos de una octava más alta, aproximadamente 220 ciclos por segundo. Cuanto más agudo sea el tono de la voz, más vibraciones se producirán (pág. 133). El grito de un niño recién nacido, alcanza como media las 400 vibraciones por segundo.

Tono. Nosotros podemos alterar la frecuencia de las vibraciones de las cuerdas vocales a voluntad, dentro de ciertos límites, para producir variaciones en el tono y en la intensidad. La utilización lingüística de estos rasgos (en relación con la «entonación», el «acento» y los «tonos» de las lenguas tonales) se describe aparte en § 29.

Oclusión glotal. También se pueden mantener las cuerdas vocales estrechamente cerradas (cuando aguantamos la respiración, por ejemplo). Cuando se abren, el aire pulmonar liberado origina una oclusión glotal [ʔ], que se escucha con bastante claridad al comenzar a toser bruscamente, pero que en muchas lenguas y dialectos se utiliza corrientemente como unidad de sonido (§ 27). En el inglés británico, por ejemplo, la oclusión glotal suele aparecer en los dialectos que muestran influencias del habla de Londres (en palabras como *bottle*, donde sustituye al sonido [t]).

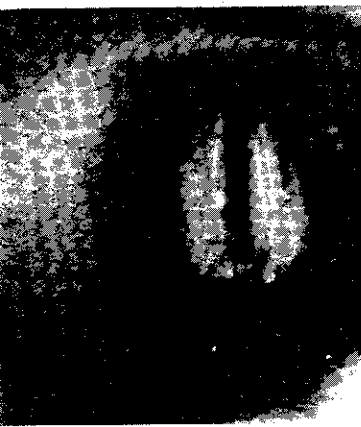
Fricción glotal. Si se mantienen las cuerdas vocales completamente separadas, el aire, expulsado energicamente, producirá el efecto de una fricción glotal audible, que las lenguas emplean a menudo como sonido [h].

Timbres de voz. Se pueden originar otros movimientos de las cuerdas vocales para producir efectos sonoros como el de la voz *susurrante* o *velada* (que podemos escuchar en las voces «de alcoba» de muchas cantantes y estrellas de cine), y la voz rechinante (por ejemplo en el amenazador tono profundo del actor de películas de terror Vincent Price), en la que las vibraciones son extremadamente lentas (aproximadamente 30 veces por segundo). Este y otros efectos similares implican esquemas de vibración muy complejos, cuyo mecanismo fisiológico aún no ha sido comprendido enteramente.



La estructura de la laringe

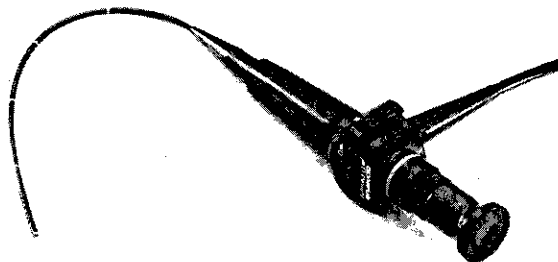
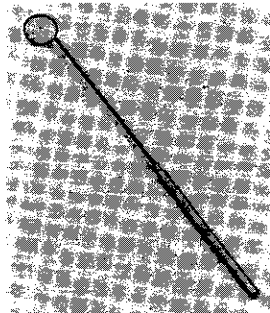
- (a) visión frontal
- (b) visión posterior
- (c) visión lateral



¿Cómo se produce la vibración de las cuerdas vocales?

La vibración de las cuerdas vocales se puede observar utilizando un espejo laringal o un laringoscopio de fibra óptica. Las vibraciones rápidas se ven como un temblor, algo así como la vibración de la cuerda de una guitarra. Para estudiar los efectos con detalle, hay que filmar la vibración y observarla a cámara lenta, algo que hoy se hace ya de forma rutinaria en la fonética instrumental (§ 24). Se han propuesto distintas teorías sobre la vibración de las cuerdas vocales. Al principio se pensaba que las cuerdas vocales vibraban de la misma manera que las de un instrumento de cuerda, como resultado directo del impulso nervioso que movía los músculos de la laringe (hipótesis que recientemente ha sido denominada teoría «neurocronáxica»). Hoy en día, en cambio, se tiende a explicar la actividad de las cuerdas vocales basándose en el modo en que los músculos de la laringe cambian la tensión y la elasticidad de las cuerdas, en respuesta a la corriente de aire que sube de los pulmones: es la teoría «mioelástica aerodinámica» de la fonación. Según este planteamiento, la presión del aire provoca la apertura de las cuerdas vocales en cada vibración; luego éstas se cierran otra vez, en parte porque su elasticidad natural hace que vuelvan a su posición anterior, y en parte porque, a causa de la brusca disminución de la presión en la glotis, cuando el aire sale por el hueco restringido de la laringe, la «succión» hace que se junten (el «efecto Bernoulli», llamado así por el matemático suizo Daniel Bernoulli (1700-82)).

El movimiento de las cuerdas vocales, filmado por una cámara de alta velocidad (10.000 imágenes por segundo), antes de que comience la fonación, con las cuerdas vocales en proceso de cerrarse (arriba), y durante la fonación, con las cuerdas abiertas y luego cerradas (abajo). Abajo, en cada foto, aparece la parte posterior de las cuerdas.



Un espejo laringal (arriba, izquierda). El espejo se introduce cuidadosamente a través de la boca hasta que se encuentra encima de la laringe. Entonces el ángulo del espejo permite una visión clara. El investigador tiene que mantener baja la lengua del sujeto, y cuidarse de que el espejo no toque las paredes de la garganta, lo que produciría un reflejo de deglución.

Un laringoscopio de fibra óptica (arriba, derecha). El tubo flexible se introduce a través de la nariz y de la cavidad nasal, hasta que cuelga por detrás del paladar blando. Algunos de los filamentos de vidrio que hay en el tubo proporcionan una potente fuente de luz, y otros envían una imagen a la pieza ocular (que puede conectarse a una cámara, pág. 141). La ventaja principal de este método, respecto al espejo laringal, es que el sujeto puede seguir hablando con el aparato colocado.

LA ARTICULACIÓN

Una vez que la corriente de aire ha pasado a través de la laringe, entra en la larga estructura tubular conocida como tracto vocal. Aquí se ve afectada por la acción de varios órganos móviles: la lengua, el paladar blando y los labios, que actúan conjuntamente para producir una amplia gama de sonidos. La producción de los diferentes sonidos del habla por medio de estos órganos recibe el nombre de *articulación*.

Además, los sonidos producidos en la laringe o tracto vocal resultan determinados por las propiedades inherentes de cada una de las cavidades por las que pasa la corriente de aire: faríngea, bucal y (a veces) nasal. Estas cavidades contagian a los sonidos su propia resonancia. Se pueden producir muchos tipos de resonancia, ya que el tracto vocal puede tomar muchas formas distintas.

Al describir la articulación, se suele establecer una distinción entre las partes inmóviles del tracto vocal («articuladores pasivos») y las que pueden moverse bajo el control del hablante («articuladores activos»). Dentro de la primera categoría, es preciso reconocer:

- los dientes superiores, especialmente los incisivos, que se utilizan para formar un estrechamiento en la articulación de algunos sonidos, como el primer sonido de *cero* [θ];
- las protuberancias situadas tras los dientes superiores, conocidas como *alveolos*, en las que se producen muchos sonidos, como [t] y [s]; y
- el arco óseo que sigue a las protuberancias alveolares, conocido como *paladar duro*, que se usa en la articulación de ciertos sonidos, como el primero de *chico* [ç].

Todos los demás órganos son móviles, en mayor o menor grado.

Articuladores activos

La faringe. Es un largo tubo muscular que va desde la cavidad laringea hasta la parte posterior de las cavidades oral y nasal. Las distintas áreas colindantes con estas cavidades proporcionan un medio para dividir la faringe en secciones: laringofaringe, orofaringe y rinofaringe. La faringe se puede estrechar o ensanchar. Mediante un estrechamiento en esta zona se pueden producir varias consonantes (pág. 155), y algunos movimientos de la laringe, del paladar blando y de la lengua pueden implicar también modificaciones faríngeas, que afectan a la cualidad del sonido. En varias lenguas (por ejemplo el árabe) existen consonantes y vocales faríngeas.

Paladar blando o velo. Es una ancha banda de tejido muscular que se encuentra en la parte posterior de la boca, cuyo rasgo más notable es la úvula, un apéndice que cuelga al fondo de la cavidad oral, fácilmente visible con la ayuda de un espejo.

En la respiración normal, el paladar blando permanece bajo, para permitir el paso del aire a través de la nariz, si bien, por supuesto, la boca puede estar abierta también. Durante el habla, existen tres posiciones principales que afectan a la cualidad de los sonidos:

- (i) El paladar blando puede permanecer alzado contra la rinofaringe para formar un «cierre velofaríngeo», de manera que el aire pueda salir sólo por la boca. Esto produce una gama de sonidos orales, como las vocales y la mayoría de las consonantes del español.
- (ii) El paladar blando puede permanecer bajo, para permitir que el aire salga por la boca y la nariz. Esta es la posición necesaria para producir *vocales nasalizadas*, como las del francés (por ejemplo en *bon* «bueno»), el portugués y muchas otras lenguas.
- (iii) El paladar blando puede permanecer bajo, al tiempo que la boca se mantiene cerrada. En este caso, todo el aire se expulsa a través de la nariz, como ocurre con las *consonantes nasales*, del tipo de [m] y [n].

Los labios. El *orbicularis oris* ('músculo que rodea la boca') es el músculo principal en el control del movimiento de los labios, aunque también participan otros músculos faciales. Los labios se pueden cerrar por completo (como para la [p] y la [m]), o se pueden abrir en varios grados para producir los distintos tipos de redondeamiento o de extensión usados en las vocales (por ejemplo [u] frente a [i], pág. 153), o la fricación de ciertos tipos de consonantes (como la *b* del español «saber»).

La mandíbula. El hueso de la mandíbula tiene un alto grado de movilidad. Controla el tamaño del hueco que queda entre los dientes e influye mucho sobre la posición de los labios. Los hablantes pueden adoptar una posición de mandíbulas abiertas o cerradas, como cuando alguien habla «apretando los dientes».



Una radiografía que muestra una visión lateral de los órganos fonadores durante la articulación de la vocal [j]. Se percibe claramente la silueta curvada de la lengua. Para un diagrama anatómico tradicional de esta zona, ver pág. 383.

La lengua

De todos los órganos móviles, la lengua es el más versátil. Es capaz de adoptar más formas y posiciones que cualquier otro órgano articulatorio, y de esta manera entra en la definición de un gran número de sonidos del habla: todas las vocales y la mayoría de las consonantes. La lengua es un músculo tridimensional, que puede moverse por entero en cualquiera de las tres direcciones principales mediante la acción de varios músculos «extrínsecos»: arriba/adelante (por ejemplo para la [i]), arriba/atrás (por ejemplo para la [u]) y abajo/atrás (por ejemplo para la [a]). Además varios músculos «intrínsecos» determinan su forma en cualquier posición. Por ejemplo, algunos músculos suben o bajan la punta de la lengua, o la mueven hacia la derecha o hacia la izquierda. Otros mueven la lengua lateralmente, o forman un surco en el medio (como se necesita para articular la [s]).

No hay secciones anatómicas de la lengua que resulten evidentes, así que, para clasificar los sonidos, es necesario hacer divisiones arbitrarias, basadas en la posición de la lengua respecto a la parte superior de la boca. Las principales zonas se localizan mejor cuando la lengua está en reposo, con la punta detrás de los dientes inferiores.

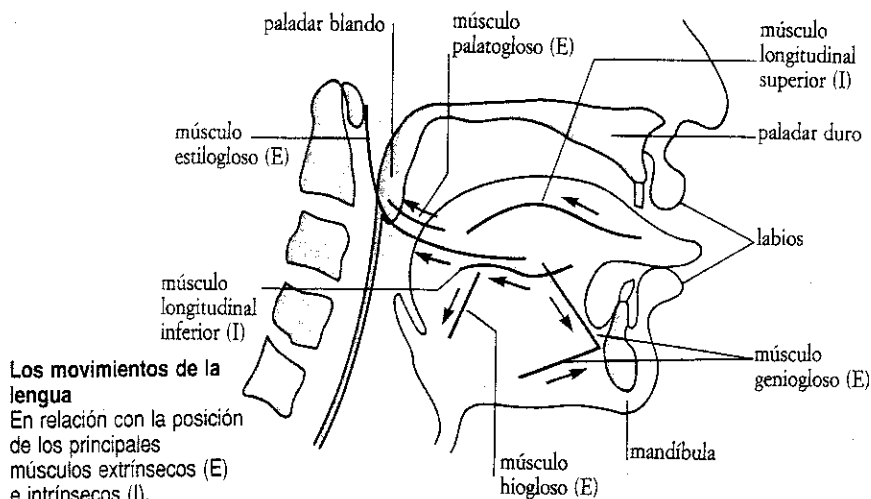
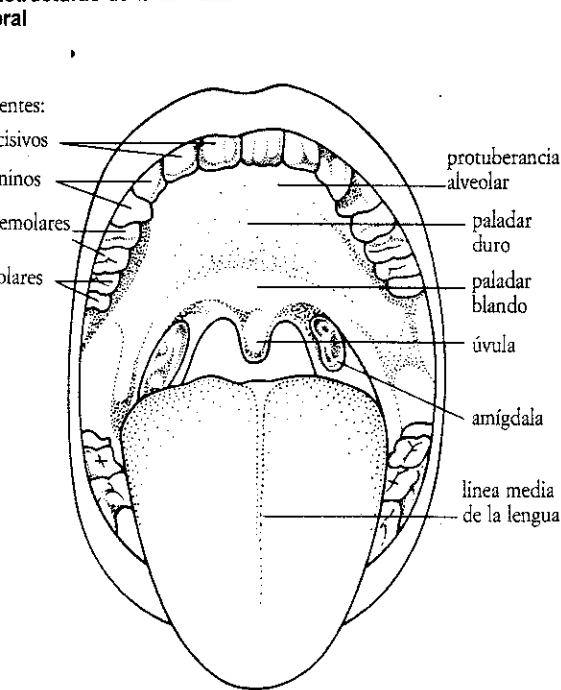
- *predorso* la parte situada bajo el paladar duro
- *postdorso* la parte situada bajo el paladar blando
- *medio dorso* la parte central.

(Estas tres partes conjuntamente constituyen el *dorso* de la lengua)

- *corona* la parte ahusada de la lengua, frente a los dientes
- *punta* o *ápice* la extremidad anterior
- *bordes*

(Las variaciones en tamaño de la lengua se comentan en § 6.)

Las estructuras de la cavidad oral



Los movimientos de la lengua
En relación con la posición de los principales músculos extrínsecos (E) e intrínsecos (I).

Los nervios que controlan los órganos articulatorios

Hay doce nervios *craneales* cuya tarea es la de conectar el cerebro con la cabeza y el cuello. Algunos cumplen una función motriz, controlando la acción de los músculos, otros cumplen una función «sensorial», enviando señales al cerebro. (Las principales zonas del cerebro que participan en la producción y percepción del habla las presentamos en § 45). Siete de los nervios craneales entran en funcionamiento como parte del proceso de habla y audición; a continuación damos una lista de sus funciones relativas. (Es corriente usar números romanos cuando se habla de los nervios craneales.)

V El nervio *trigémino* actúa como motor de los músculos de la mandíbula y de uno de los músculos que controlan el paladar blando. También actúa como nervio sensorial desde los dos tercios posteriores de la lengua.

VII El nervio *facial* es un nervio motor que actúa sobre los músculos de los labios.

VIII El nervio *auditivo* o *acústico* actúa como nervio sensorial desde el oído (pág. 143).

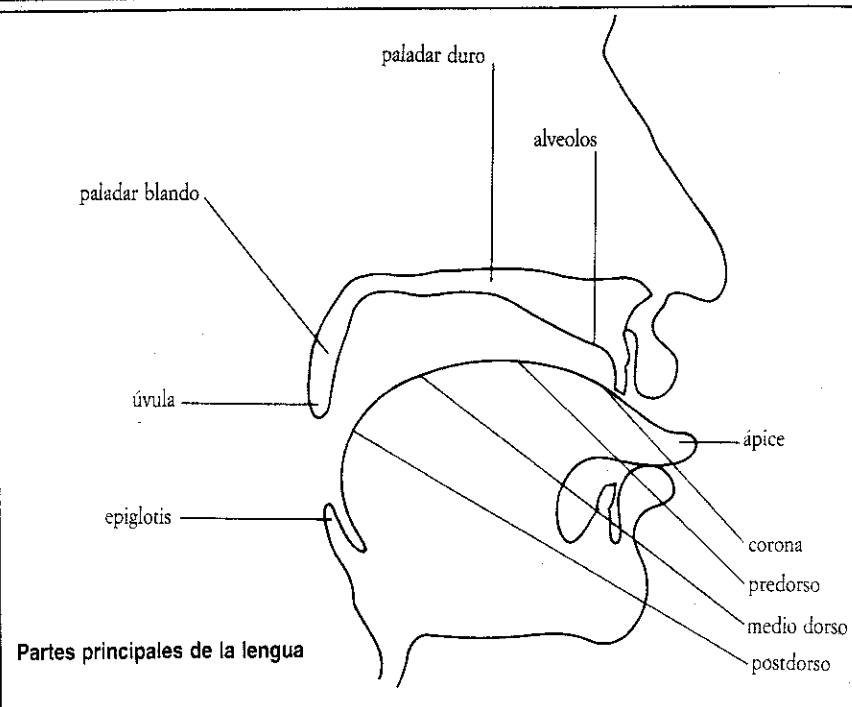
IX El nervio *glosofaríngeo* actúa sea como nervio motor para la faringe, sea como nervio sensorial para el postdorso de la lengua.

X El nervio *vago* es un nervio motor que actúa sobre los músculos de la laringe y la faringe.

XI El nervio *accesorio/espinal* actúa como motor para el alzamiento del paladar blando.

XII El nervio *hipogloso* es un nervio motor que actúa sobre los músculos de la lengua.

Además habría que notar la importancia de varios nervios *espinales*, algunos de los cuales controlan los músculos del pecho que participan en la respiración (pág. 124).



Partes principales de la lengua

23. La acústica del habla

La energía del sonido es una onda de presión que consiste en la vibración de las moléculas en un medio elástico, como por ejemplo un gas, un líquido o algunas clases de sólidos. Al estudiar la producción del habla, normalmente nos ocupamos de la propagación del sonido a través del aire: las partículas de aire son perturbadas por los movimientos y las vibraciones de los órganos articulatorios, especialmente por las cuerdas vocales (§22). Pero cuando estudiamos la recepción del habla (§25), el aire no es el único medio implicado. El proceso de la audición requiere que las vibraciones en el aire se transformen en vibraciones mecánicas (a través de los mecanismos óseos del oído medio), cambios hidráulicos (a través de los líquidos del oído interno), e impulsos eléctricos nerviosos (por el nervio auditivo hasta el cerebro).

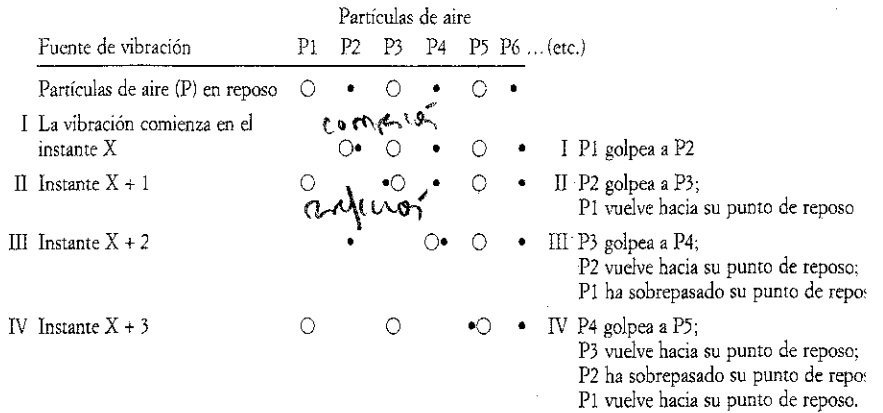
Cuando un objeto vibra, provoca un movimiento oscilante en las partículas de aire que lo rodean. Estas partículas afectan a partículas adyacentes, y el proceso continúa como una reacción en cadena mientras dura la energía. Si la vibración inicial conlleva una gran cantidad de energía, la onda sonora producida puede transmitirse a gran distancia, antes de desaparecer. Pero las propias partículas de aire no viajan a lo largo de toda esta distancia. El movimiento de cada partícula es meramente local: cada una influye en la siguiente, de la misma manera que una larga fila de fichas de dominó, colocadas unas cerca de otras, se derrumba si empujamos la primera. No obstante, a diferencia de las fichas de dominó, las partículas de aire vuelven a su posición inicial, una vez que han transmitido el movimiento a sus vecinas.

El movimiento de las ondas sonoras en el aire se ha comparado a veces con las olas que provoca una piedra cuando cae al agua; pero esta analogía no refleja la naturaleza tridimensional del movimiento. De la misma manera, el símil del dominó sólo da una idea limitada del movimiento que se produce. Una comparación mejor sería la de un globo que se infla, que se expande en todas las direcciones a la vez. También las ondas sonoras se mueven simultáneamente en todas las direcciones, desde el punto de origen.

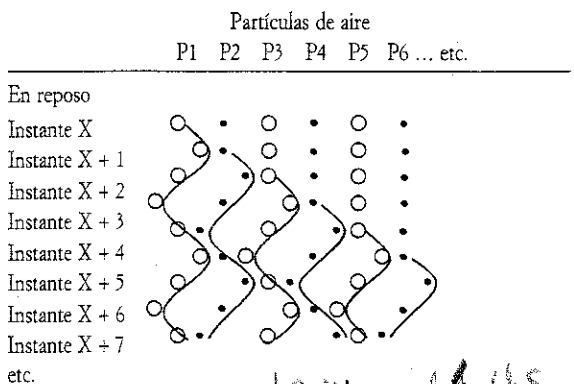
SINUSOIDES

El movimiento de las partículas puede compararse al de un péndulo o un columpio. Cuando está parado, el columpio cuelga verticalmente. Cuando se pone en movimiento, a un desplazamiento hacia atrás le sigue uno hacia adelante, a ambos lados del punto de reposo, mientras haya energía para mover el columpio. Este vaivén se conoce como *oscilación*. De igual manera, las partículas de aire oscilan alrededor de su punto de reposo. Cuando una partícula se mueve hacia adelante comprime a las partículas adyacentes, y causa un ligero aumento de la presión del aire en ese punto. Cuando se mueve hacia atrás deja de comprimir a esas partículas, y provoca una disminución de la presión. El movimiento es como el de una ola,

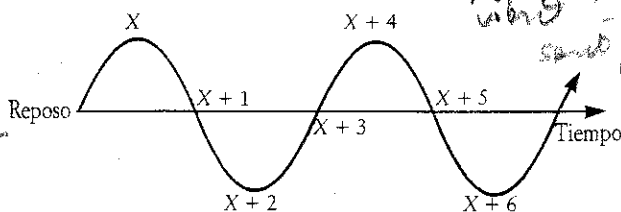
como podemos observar si seguimos el progreso de una serie de partículas, una vez que han sido puestas en movimiento por una fuente de vibraciones. En el siguiente diagrama, se representa el movimiento de cada partícula en instantes sucesivos (X + 1, X + 2, etc.), suponiendo que la fuente de vibración esté a la izquierda.



Si unimos con una línea las posiciones de cada partícula, podremos apreciar el movimiento ondulatorio, que seguidamente presentamos en vertical.



Se puede trazar un gráfico de la onda de presión que se forma cuando las partículas se mueven de esta manera; es lo que se conoce como *sinusoide*. Las ondas sinusoidales se suelen trazar de izquierda a derecha, a ambos lados de una línea horizontal que representa el paso del tiempo. El movimiento simple de una sola partícula tendría este aspecto:



Sonidos puros

Las ondas más sencillas, como las que aparecen en esta página, tienen forma de seno, y consisten en una sola pulsación vibratoria, que se repite a intervalos constantes y produce un *sonido puro*. Estos sonidos difícilmente se escuchan en la vida diaria. La mayoría de los sonidos son compuestos, al estar formado por varios patrones simultáneos de vibración. Par producir un sonido puro, hace falta una máquina electrónica especial, o un instrumento como el diapasón. Cuando se golpea un diapasón (abajo), éste vibra sobre una sola nota. Las puntas de la horquilla oscilan con un ritmo constante. Si se acerca el diapasón al oído, se puede percibir un sonido puro



propiedad de la onda

A la oscilación de una partícula de aire se la llama *ciclo*, al número de ciclos que se suceden en un segundo se le llama *frecuencia* de un sonido. La frecuencia se solía medir en «ciclos por segundo» (cps), pero a esta unidad se le ha dado el nombre de *hertzio* (por el físico alemán Heinrich Rudolf Hertz (1857-94), que fue el primero en emitir y recibir ondas de radio), abreviado en Hz. La frecuencia básica de la vibración de un sonido se conoce como *frecuencia fundamental*, abreviado generalmente en F0, pronunciado «F cero».

La gama de frecuencias que puede oír un adulto joven normal es amplísima, desde unos 20 a 20.000 Hz. No se pueden oír las frecuencias que están por debajo («infrasónicas») o por encima («ultrasónicas») de estos valores. En cualquier caso, las frecuencias que se encuentran a ambos extremos de esta gama tienen poco interés para el habla: las frecuencias de habla más importantes se encuentran entre los 100 y los 4.000 Hz. La frecuencia fundamental de la voz de un hombre adulto, por ejemplo, es de unos 120 Hz, y la de una mujer de unos 220 Hz (pág. 128).

La frecuencia de un sonido puro guarda relación con la sensación de tono, que es nuestra percepción de que un sonido es «más alto» o «más bajo». En general, cuanto más alta es la frecuencia de un sonido, más alto es el tono que percibimos. Pero nuestra percepción del tono también se ve afectada por la duración y la intensidad del estímulo sonoro. Las nociones de «frecuencia» y «tono» no son idénticas: la frecuencia es un hecho físico, objetivo, mientras que el tono es una sensación psicológica, subjetiva (pág. 144).

Longitud de onda

La *velocidad* a la que el sonido viaja a través del aire tiene un *valor constante*, que habitualmente (según las condiciones de temperatura) corresponde a unos 343 m. por segundo. Todos los sonidos que tienen la misma energía tardarán el mismo tiempo en desplazarse de A a B. Durante el tiempo necesario para que se produzca un ciclo de la vibración, una onda sonora recorre una cierta distancia, que constituye la *longitud de onda* del sonido. Puesto que la velocidad a la que se mueve el sonido es constante, resulta que *cuanto más alta sea la frecuencia, más corta será la longitud de onda*. Dicha relación se expresa con una sencilla fórmula: $\lambda = (C/F)$, donde C es la velocidad del sonido, F es la frecuencia, y lambda es la longitud de onda. De este modo, un sonido de 500 Hz tendrá una longitud de onda de $(343 \text{ m.} / 500 \text{ Hz}) = 69 \text{ cm.}$; la de un sonido de 1.000 Hz sería de 34 cm.

La importancia de la longitud de onda se aprecia en nuestro modo de recibir el sonido: cuando una onda se aproxima a un objeto, si su longitud de onda supera el tamaño del objeto, tiende a «combarse» alrededor de éste, mientras que si la longitud de onda es menor, tiende a rebotar. Así, por ejemplo, cuando las ondas sonoras se acercan a la cabeza, es más probable que sean retenidas las frecuencias más bajas, que tienen una longitud de onda mayor, antes que las altas, factor que puede tener una importancia considerable cuando se trata de asistir a personas con problemas de audición (pág. 266).

La mayoría de las fuentes de sonido producen grupos complejos de vibraciones, y ese es siempre el caso del habla. El habla implica la utilización de ondas sonoras complejas porque es el resultado del empleo de muchas fuentes distintas de vibración dentro del tracto vocal (§ 22). Cuando se combinan dos o más ondas puras de frecuencia distinta, el resultado es una *onda compleja*. Existen dos tipos de onda compleja. En uno de ellos la onda sonora se repite: hay un patrón de vibración *periódico*. En el otro no hay tal repetición: las vibraciones se producen al azar, de forma *aperiódica*. En el habla se emplean ambos tipos. Los sonidos vocálicos, por ejemplo, presentan un patrón periódico, mientras los sonidos como el de la [s] son aperiódicos.

Armónicos

de resonancia

El sonido producido por un objeto que vibra de forma periódica implica más que una simple senoide (pág. 132). La misma vibración genera también otras cantidades de energía, ligadas a la senoide básica según una sencilla relación matemática: *todas son múltiplos de la frecuencia fundamental*. Es decir, una F0 de 200 Hz originaría una serie de frecuencias relacionadas de 400 Hz, 600 Hz, etc. Estos múltiplos se conocen como *armónicos* o *hipertonos*, y se numeran por orden. En física (pero no en música) F0 cuenta como el primer armónico, o sea que en el presente ejemplo 400 Hz sería el «segundo armónico», 600 Hz el «tercer armónico», etc. Este sistema constituye un marco adecuado para el estudio de las vocales, de algunas consonantes, y de los patrones de entonación (págs. 135-7).

Según cuál sea la naturaleza del objeto que vibra (por ejemplo el material de que está hecho, o su espesor), se establecen grupos de armónicos diferentes, que nosotros percibimos como diferencias de *timbre*. La diferencia que apreciamos entre dos voces o dos instrumentos musicales que producen un sonido con el mismo tono y la misma intensidad, es un contraste en el timbre, causado por la diferencia de los armónicos.

Frecuencias familiares

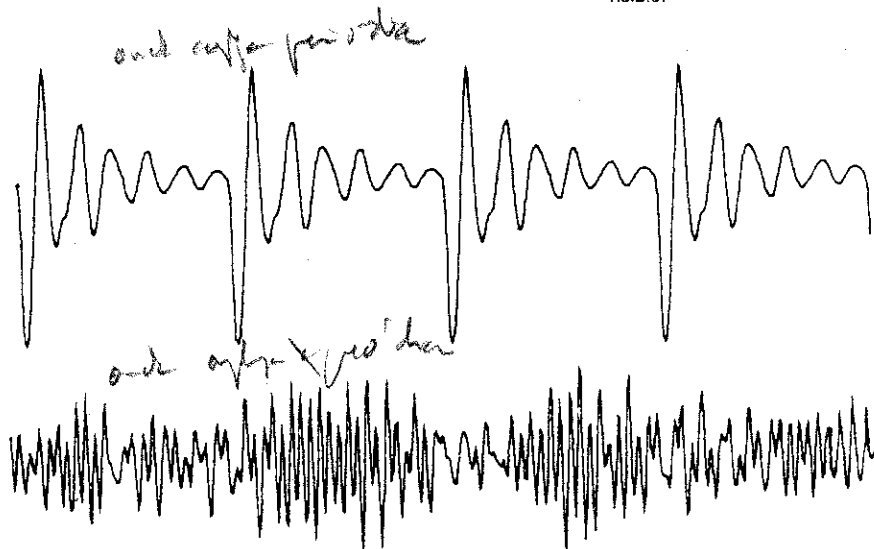
Una manera de relacionar la noción física de frecuencia con nuestra percepción del tono consiste en relacionar las notas musicales que nos resultan familiares con la frecuencia fundamental. El Do mayor tiene una frecuencia de 264 Hz, y la frecuencia de las notas que están por encima de éste en la escala diatónica de Do son las siguientes:

Do	528 Hz
Si	495 Hz
La	440 Hz
Sol	396 Hz
Fa	352 Hz
Mi	330 Hz
Re	297 Hz
Do	264 Hz

La nota La es la nota que toca el oboe cuando la orquesta está afinando. En comparación, la nota más alta de un piano de siete octavas es de 3.520 Hz, y la más baja de 27,5 Hz.

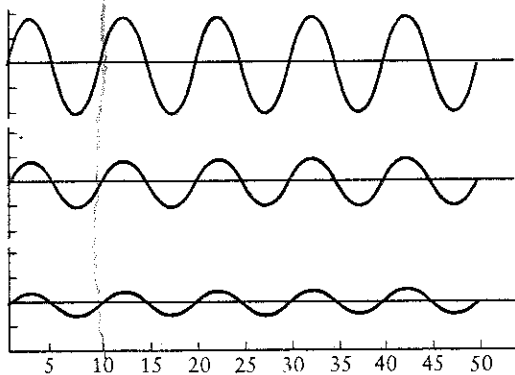
Ondas vocales

He aquí las ondas sinusoidales típicas de la vocal [a:] y de la consonante [s]. El espacio de tiempo que se representa es el mismo para ambos sonidos. En el caso del sonido vocálico el patrón periódico se puede apreciar claramente, pero en el caso de la [s] no hay ningún patrón visible.



AMPLITUD E INTENSIDAD

La distancia alcanzada por una partícula de aire a un lado y a otro del punto de reposo se conoce como **amplitud** de la vibración. Cuanto mayor es la amplitud, mayor es la **intensidad del sonido** y, por efecto también de otros factores (como la frecuencia y la duración), mayor es la sensación de altura que percibimos. En el siguiente diagrama podemos ver tres ondas de igual frecuencia pero distinta amplitud. En todos los casos una vibración completa dura 10 msec (por lo tanto la frecuencia es de 100 por segundo, o 100 Hz). Pero la amplitud de (a) es el doble que la de (b), y la de (b) el doble que la de (c).



Para medir la altura de un sonido tenemos que tener en cuenta tanto la **amplitud** como la **frecuencia**, dos factores que están relacionados con la energía con que el sonido se produce. El término **intensidad** se emplea para referirse a la potencia global de un sonido, y es una noción muy útil para el estudio del habla, donde las ondas sonoras son complejas, y la altura de un sonido no se relaciona claramente con ninguno de los componentes acústicos.

Decibelios

Para medir la intensidad del sonido necesitamos un nivel básico de referencia de la presión del sonido en el aire, que esté reconocido internacionalmente. Este nivel de presión del sonido tomado como referencia indica el límite a partir del cual un sonido puede ser oído (que se define tradicionalmente como 0.0002 dinas por centímetro cuadrado, donde *dina* es la unidad de medida de la presión). Los valores por encima de este nivel de referencia se miden en unidades conocidas como **decibelios** (db) (llamadas así por Alexander Graham Bell (1847-1922), in-

ventor del teléfono). Por consiguiente, si decimos que un sonido es de 90 db, significa que tiene una intensidad de 90 db por encima del nivel de referencia.

Nosotros somos capaces de oír una vasta gama de intensidades de sonidos. Un grito es un millón de veces más potente que un susurro. Se ha calculado que el oído humano es sensible a unos 10 millones de millones (10 elevado a 13) de unidades de intensidad. Para que los analistas puedan manejar cifras tan grandes, las intensidades de los sonidos se relacionan entre ellas de forma proporcional, utilizando una tabla de logaritmos. Un aumento de 10 db equivale a multiplicar la intensidad por dos: 30 db es el doble de alto que 20 db, 40 db es el doble de alto que 30 db, etc. De esta manera, 10 elevado a 13 puede reducirse a una escala de 130 decibelios, escala que refleja con más exactitud el modo en que percibimos la diferencia de intensidad entre los sonidos.

Se pueden calcular los valores medios de intensidad de cada uno de los sonidos del habla. En la siguiente tabla (según D. B. Fry, 1979, pág. 127), los valores de los sonidos ingleses, expresados en decibelios, se han puesto en relación con el sonido de menor intensidad, [θ] (como en inglés *thin*), al que se da el valor 0. Las vocales abiertas son los sonidos más intensos, seguidas por las vocales cerradas; las fricativas y oclusivas débiles se encuentran en el extremo opuesto de la lista (§ 27; para las convenciones de transcripción, ver apéndice 2). De acuerdo con esto, en una palabra como *thorn* el aumento de intensidad entre el primer sonido y el segundo es de unos 30 db.

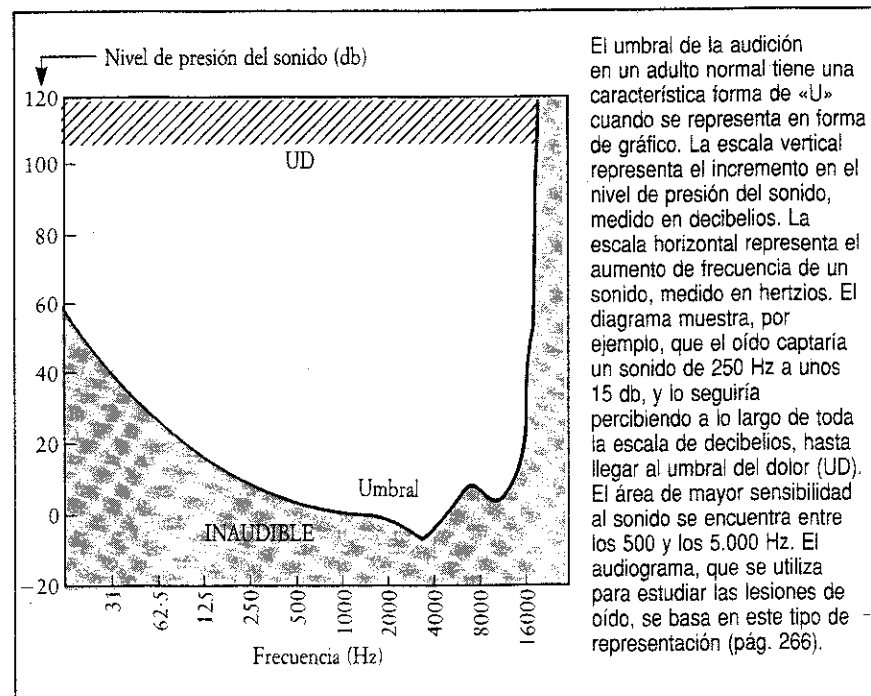
ɔ	29	e	23	l	20	ʒ	13	ð	10
ɒ	28	i:	22	ʃ	19	z	12	b	8
ɑ:	26	u:	22	ŋ	18	s	12	d	8
ʌ	26	ɪ	22	m	17	t	11	p	7
ɜ:	25	w	21	tʃ	16	g	11	f	7
a	24	r	20	n	15	k	11	θ	0
ʊ	24	j	20	dʒ	13	v	10		

Habla y sonidos cotidianos

La intensidad propia de las distintas clases de habla se puede apreciar comparándola con la intensidad media de algunos sonidos corrientes (según D. B. Fry, 1979).

0	umbral de audición
10	susurrar de las hojas
20	tictac de un reloj (pegado al oído);
30	jardín silencioso; conversación susurrada
40	zona residencial, sin tráfico
50	oficina tranquila, máquina de escribir
60	conversación a 1 m.; coche a 10 m.
70	tráfico urbano muy intenso a 30 m.
75	timbre del teléfono a 3 m.; gritos
80	tren del metro; música alta en la radio
90	taladradora a 1 m.
100	claxon de coche a 5 m.; fortissimo orquestal
110	taller de calderería
120	banda de rock con amplificación potente
130	avión de reacción de cuatro motores a 30 m.

Alrededor de los 120 db, la sensación de audición se convierte en sensación de dolor.



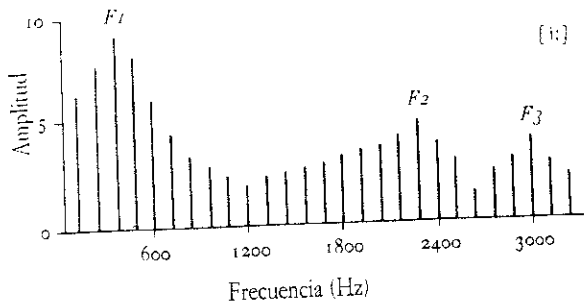
El umbral de la audición en un adulto normal tiene una característica forma de «U» cuando se representa en forma de gráfico. La escala vertical representa el incremento en el nivel de presión del sonido, medido en decibelios. La escala horizontal representa el aumento de frecuencia de un sonido, medido en hertzios. El diagrama muestra, por ejemplo, que el oído captaría un sonido de 250 Hz a unos 15 db, y lo seguiría percibiendo a lo largo de toda la escala de decibelios, hasta llegar al umbral del dolor (UD). El área de mayor sensibilidad al sonido se encuentra entre los 500 y los 5.000 Hz. El audiograma, que se utiliza para estudiar las lesiones de oído, se basa en este tipo de representación (pág. 266).

El sonido ambiental

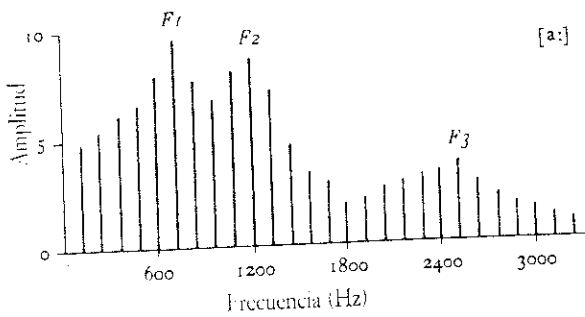
El tiempo que tardan las ondas sonoras en desvanecerse en el nivel de lo inaudible se conoce como «tiempo de reverberación». En una habitación, las paredes y los muebles absorben la energía. Los materiales fibrosos, como las cortinas y las alfombras, absorben bien el sonido, mientras que las superficies duras y densas lo reflejan. En las aulas modernas se suelen usar estos últimos materiales, que producen altos niveles de ruido («ruido ambiental»), y a menudo son la causa de que sea difícil para los niños oír lo que se dice.

ESPECTROS

Se puede hacer un análisis acústico de una onda compuesta y presentar sus distintos componentes en forma de un *espectro* de sonido. Un análisis de espectro es un gráfico en el que el eje horizontal representa la frecuencia y el eje vertical representa la amplitud. Por ejemplo, los numerosos componentes acústicos de la vocal [i:], que representan la resonancia producida en el tracto vocal durante la articulación, se pueden mostrar en forma de *espectrograma* de la siguiente manera:



La vocal [a:] proporciona un espectro distinto, lo que refleja una configuración completamente diferente del tracto vocal.



Nótese que, en estos espectros, la amplitud de algunas frecuencias es mucho mayor que la de otras. De hecho, en cada caso se pueden ver varios «picos» de energía acústica, que reproducen los principales puntos de resonancia del tracto vocal. Dichos picos se denominan *formantes*, y se numeran del más bajo al más alto: «primer formante» (F₁), «segundo formante» (F₂), etc. En el espectro de [i:], arriba (proferido por un hombre con una frecuencia fundamental de 120 Hz), F₁ alcanza 360 Hz, F₂ 2.280 Hz, y F₃ 3.000 Hz.

La estructura de los formantes es un rasgo fundamental de los sonidos del habla. Todas las vocales y algunas de las consonantes tienen formantes. El patrón de los formantes (especialmente la disposición de los dos primeros) es lo que nos permite distinguir las vocales, o reconocer que se está repitiendo la «misma» vocal, incluso cuando la pronuncian hablantes distintos. Además, los formantes de las vocales también pueden ayudar a identificar la naturaleza de los sonidos consonánticos contiguos.

La relación entre acústica y articulación

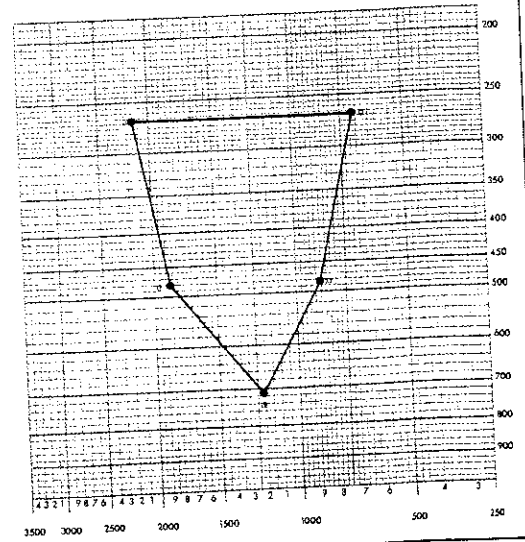
Estudiando a distintos hablantes, es posible calcular las principales frecuencias del primer y del segundo formante de las vocales de una lengua. Así se ha hecho también para el español, con los siguientes resultados (según E. Martínez Celdrán, 1984):

	F ₁	F ₂
[i]	267	2112
[e]	489	1889
[a]	711	1222
[o]	489	889
[u]	267	711

Podemos situar estas cifras en un gráfico en el que en el eje vertical aparece la frecuencia de F₁, y en el eje horizontal la de F₂. El patrón resultante es muy similar al que aparece cuando se describen las vocales de acuerdo con el punto de

articulación en el tracto vocal (véase el «trapecio» vocálico, pág. 154). No obstante, no coinciden de forma exacta, porque el esquema articulatorio se basa sólo en

el punto donde la oclusión de la lengua es mayor, mientras que el esquema acústico deriva de la resonancia del tracto vocal entero.



Analizador espectral de tiempo real. Este instrumento utiliza un tubo de rayos catódicos para mostrar el espectro en constante cambio de las ondas compuestas. Debido a la velocidad con que se suceden los cambios, para un análisis más profundo es necesario que el investigador «pare la acción», por ejemplo conservando la onda en un osciloscopio de almacenamiento, fotografiando la pantalla, o utilizando la pantalla de un ordenador.

EL ESPECTRÓGRAFO

Durante la década de 1940, se diseñó el espectrógrafo para analizar y presentar los espectros acústicos. Esta máquina graba la voz, analiza las diferentes frecuencias de las ondas sonoras mediante una serie de filtros electrónicos, mide la intensidad de cada frecuencia, y luego presenta el resultado de forma visual, utilizando una aguja para trazar las marcas sobre una tira de papel satinado especial. En esta página y en la siguiente presentamos una serie de espectrogramas que representan varios sonidos.

En esta clase de espectrogramas se muestran tres dimensiones del sonido.

1. El tiempo se representa horizontalmente; en el espectrógrafo más usado (el Sona-Graph Kay), se pueden grabar 2,4 segundos de habla sobre el papel, que se lee de izquierda a derecha. Cada media pulgada de papel (1,27 cm.) representa una décima de segundo de habla.
2. La dimensión vertical ofrece información sobre la frecuencia, desde 0 Hz (la línea inferior) hasta 8.000 Hz. Es una escala lineal: cada pulgada en vertical representa 2.000 Hz.
3. La tercera dimensión es la de la intensidad, indicada por el grado de oscuridad de las marcas sobre el papel. Cuanto más intensa es la señal, más oscuras son las marcas que deja la aguja. De esta manera, las frecuencias de intensidad muy baja o nula aparecen como zonas de papel blanco.

Si bien las limitaciones de esta máquina no permiten efectuar mediciones perfectas sobre el papel, el despliegue visual hace posible un reconocimiento inmediato para todo el que esté acostumbrado al análisis acústico, lo que ha convertido al espectrógrafo en un instrumento de investigación habitual. Se puede mostrar claramente el contraste entre sonidos distintos, al igual que la influencia de unos sonidos sobre otros cuando se mezclan en la cadena hablada. Además, en la actualidad se están haciendo grandes progresos en el diseño espectrográfico que permiten obtener análisis más exactos, gracias a los métodos informáticos. En particular, ahora se dispone de espectrógrafos digitales que pueden almacenar dos espectrogramas en pantalla, lo que permite una comparación más sistemática y más precisa de los detalles.

Tipos de espectrograma

Cuando se hace un espectrograma se puede elegir entre dos posibilidades. Si se elige la opción «estrecho», la máquina analizará la gama de frecuencias del habla en bandas pequeñas (normalmente de 45 Hz), lo que hace que los armónicos destaquen con gran claridad. Si la máquina se coloca en «ancho», el análisis se hará utilizando bandas de

frecuencia mucho más amplias (normalmente de 300 Hz), lo que hace que destaquen claramente los formantes. El análisis de banda ancha es más útil para los objetivos de la fonética, en la mayoría de los casos. Con el espectrógrafo también se puede obtener información sobre la amplitud de un sonido. La intensidad de cada componente de la frecuencia es analizada e

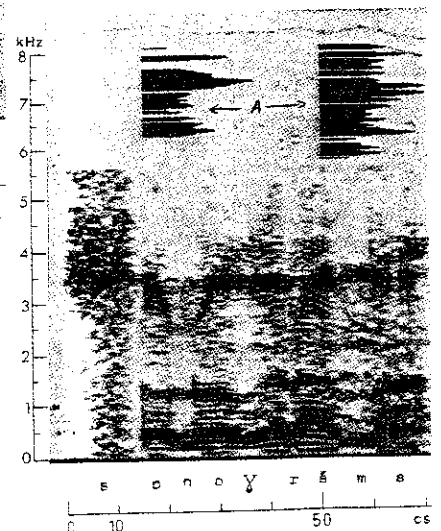
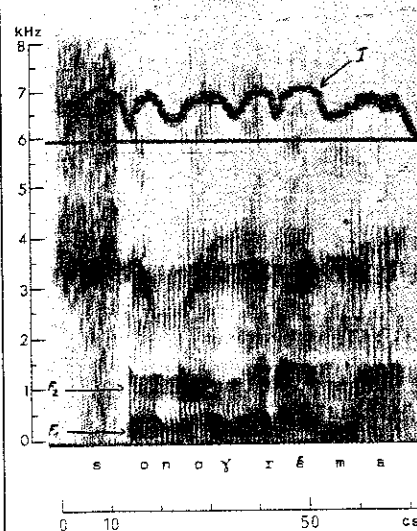
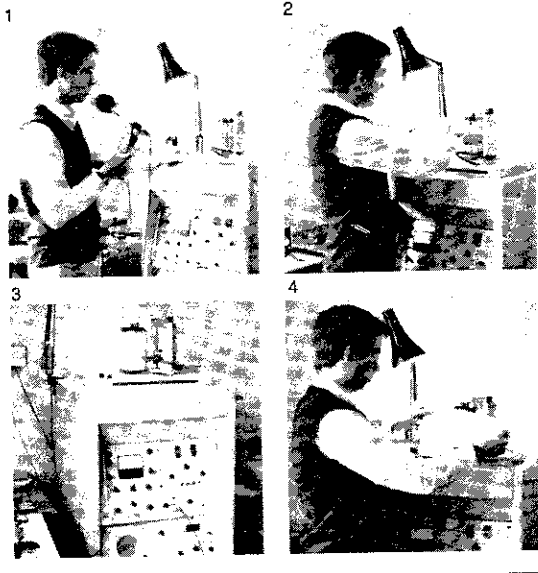
impresa en la parte superior del espectrograma, obteniéndose así una imagen de la amplitud. Abajo aparecen las versiones de banda estrecha y de banda ancha de la palabra «sonograma», la primera acompañada de una imagen de la intensidad, y la segunda de una imagen de la amplitud (según A. Quilis, 1981).

Cómo se hace un espectrograma

1. Se graban unas palabras en el espectrógrafo, bien mediante un micrófono, bien desde un magnetofón.
2. Se coloca un papel satinado especial alrededor del tambor que se encuentra sobre la máquina, y se pone una aguja en contacto con el papel.
3. Cuando se pone en

marcha la máquina, el tambor gira. Entonces la aguja sube por el tambor, analizando las distintas frecuencias del habla. Si en una frecuencia determinada hay energía, la aguja deja una marca en el papel.

4. Cuando la aguja llega al final del tambor, se desconecta la máquina y se quita el papel. El espectrograma está listo para ser interpretado.



Los rasgos acústicos de las vocales y las consonantes

Vocales Todos los sonidos vocálicos en el habla normal presentan dos o, normalmente, tres formantes, que en un espectrograma de banda ancha aparecen como espesas bandas negras. Se pueden apreciar claramente en los espectrogramas de las vocales largas [i:], [a:] y [u:], pronunciadas aisladamente. Las estrias verticales representan las vibraciones de las cuerdas vocales.

Semivocales y semiconsonantes. Los sonidos [j] y [w] (como en *bien* y *agua*) funcionan como consonantes en muchas lenguas, pero poseen los rasgos acústicos de las vocales [i] y [u] respectivamente; se les suele llamar «semiconsonantes» (pág. 152). Su naturaleza vocálica se puede observar claramente en un espectrograma en el que aparezcan articuladas entre dos vocales [a]. En ambos casos, los formantes se inclinan cuando la vocal cambia de timbre. Por ejemplo, en [aja] el primer formante se inclina hacia abajo y el segundo se inclina hacia arriba, cuando la lengua pasa de [a] a [j]; se vuelven a inclinar cuando la lengua vuelve a la posición original. La inclinación también afecta al tercer formante, pero afecta mucho menos al cuarto formante, que se encuentra más arriba.

Consonantes oclusivas Lo que habitualmente identifica a

una consonante oclusiva es un breve silencio, mientras en la boca se crea una oclusión, seguido de una explosión de sonido, cuando la oclusión se relaja (pág. 157). Los dos rasgos se pueden observar en un espectrograma, especialmente si dichas consonantes se articulan entre vocales. El silencio corresponde a la tira vertical de papel blanco y la relajación a una delgada «punta» de marcas, ampliamente esparcidas por el espectro. La aparición de las bandas negras de los formantes marca el comienzo de la siguiente vocal. Se puede observar claramente la diferencia entre las consonantes sordas [p, t, k] y sonoras [b, d, g] (pág. 128).

• En los sonidos sonoros la intensidad de la explosión de sonido es mucho menor que en los sonidos sordos.

• La duración del silencio es menor en los sonidos sonoros.

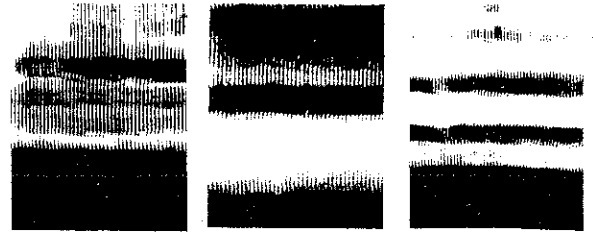
• Una variable importante es el tiempo que transcurre entre la relajación de la oclusiva y el comienzo de la vibración de las cuerdas vocales (representado por los formantes). Esta variable se conoce como «tiempo del comienzo de la voz» o «voice onset time» (VOT). Existe un hueco de VOT considerable en el caso de las oclusivas sordas (indicado en los espectrogramas con una X) mientras que en las oclusivas sonoras el hueco

es muy pequeño o inexistente. En tales casos, la sonoridad puede comenzar antes que la explosión de sonido.

En los espectrogramas también se pueden ver los distintos puntos de articulación de las consonantes oclusivas. Los sonidos bilabiales [p, b] presentan una explosión sonora en las frecuencias bajas; los sonidos alveolares [t, d] muestran dicha explosión en las frecuencias altas, y los sonidos velares [k, g] la tienen en la zona intermedia. También hay una clara diferencia en el punto de transición entre consonante y vocal: en dicho punto se produce un rápido movimiento articulatorio, que se refleja en la brusca inclinación de los formantes donde comienza la vocal; inclinación que cambia de dirección entre cada par de consonantes.

Consonantes fricativas Estas consonantes utilizan una energía acústica casual, o ruido, representada en el espectrograma por una ancha zona de perturbación en ciertas frecuencias. Donde se aprecia con más claridad es en el caso de las sibilantes fricativas, como [s] y [ʃ], que son sonidos con una energía elevada. La energía de la [s] está muy por encima de los 4.000 Hz, mientras que la de [ʃ] empieza más abajo, alrededor de los 2.500 Hz. Aquí presentamos el contraste de estos sonidos entre vocales [a].

Vocales

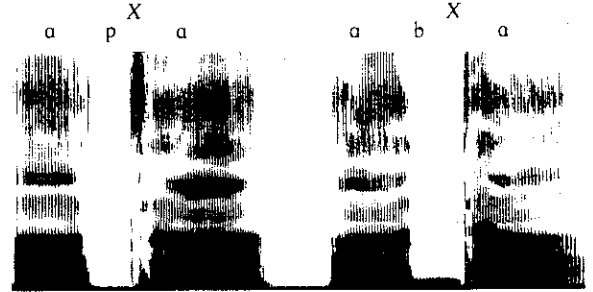


a: i: u:

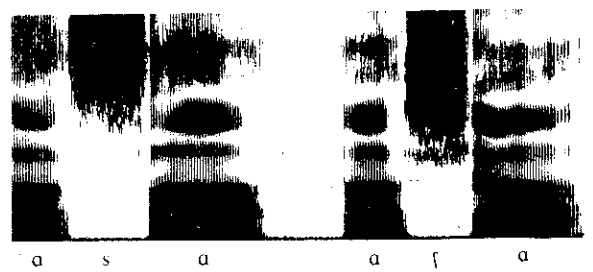
Semivocales y semiconsonantes



Consonantes oclusivas



Consonantes fricativas

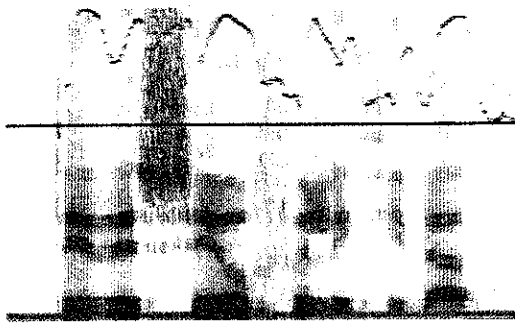


La cadena hablada

Los espectrogramas de la derecha corresponden a unidades de sonido articuladas con cuidado, de forma aislada. Pero en la cadena hablada, los sonidos se producen con mayor rapidez, influyen unos en

otros, y resulta más difícil distinguir los límites que los separan. Los espectrogramas de la cadena hablada ponen el acento en la continuidad fundamental de la lengua hablada, en el hecho de que la articulación es un proceso

de cambio continuo. Es algo que siempre habría que tener en cuenta cuando se leen los símbolos, claros y aislados, de la transcripción fonética. (El ejemplo es de Martínez Celdrán, 1984, pág. 327).



«Te deseo felicidad»

24. El análisis instrumental del habla

Hoy en día se dispone de una vasta gama de instrumentos para el estudio de la producción del habla. Un laboratorio de fonética bien equipado posee equipos para grabar el habla, analizar las propiedades acústicas de la señal sonora (§23), e investigar la fisiología de los órganos vocales (§22). Asimismo disponemos de las técnicas correspondientes para el estudio de la audición y de la recepción del sonido (§25). Este campo se conoce globalmente como *fonética instrumental* (o *experimental*).

GRABAR EL HABLA

Cualquier investigación científica de la naturaleza del habla requiere la conservación de grabaciones permanentes. La señal acústica puede grabarse en un disco o (más corrientemente) en una cinta de audio, y presentarse de forma visual en un gráfico sobre papel, en una impresión por ordenador o en una pantalla (que luego se puede fotografiar). De la misma manera, se dispone de la fotografía y de otras técnicas para registrar y presentar la actividad fisiológica del hablante.

La cuidadosa elección de los métodos e instrumentos tiene la más alta prioridad en la investigación acústica. A menos que se tomen precauciones especiales, las grabaciones podrían carecer de la claridad necesaria para permitir que se realice un análisis acústico preciso. Si una grabación contiene mucho ruido de fondo, o si la señal es débil o está distorsionada, las ondas sonoras del habla serán confusas. Por lo tanto hay que prestar especial atención a las limitaciones del instrumento de grabación (normalmente un magnetofón), al micrófono, al sistema de reproducción y al lugar en que se realiza la grabación.

Para obtener resultados óptimos, las grabaciones deberían efectuarse en un estudio especial, aislado de los sonidos del exterior y dotado de paredes que absorban el sonido. Si no se dispone de esto, habría que efectuar las grabaciones en una habitación en silencio que contenga material que absorba el sonido (como muebles de materiales blandos). De esta manera, se reducirían al mínimo los problemas de eco y otras interferencias.

MOSTRAR LAS SEÑALES DEL HABLA

El instrumento que se usa más habitualmente para observar las ondas sonoras es el *osciloscopio*, que muestra la frecuencia y la amplitud de una onda sinusoidal. Algunos instrumentos sólo ofrecen una imagen temporal; otros (los osciloscopios de almacenamiento) pueden mantener la imagen de una onda en la pantalla para un estudio más detallado. En estos casos, las imágenes se pueden fotografiar directamente, o se puede utilizar algún tipo de registrador gráfico para obtener una señal sobre papel.

Existen muchas clases de reproductores gráficos, que reflejan distintas tecnologías. Todos implican el uso de una tira de papel que se desenrolla mecánicamente, sobre la que se puede trazar una señal mediante una o más plumas o chorros de tinta (según cuántos canales se graben a la vez), o el uso de un sistema basado en la grabación por calor, fibras ópticas o luz ultravioleta. Ahora ya se dispone habitualmente de varias formas de procesamiento por ordenador de las señales del habla, en las que la información se puede presentar rápidamente en forma gráfica o numérica.

El espectrógrafo, el más valioso de todos los instrumentos por su presentación visual del habla, aparece en la página 136.



Una cámara anecoica (arriba). El suelo, el techo y las paredes han sido recubiertos de un material diseñado para absorber el sonido y eliminar la reverberación.

Un osciloscopio de almacenamiento (izquierda).

Un Visicorder (abajo). En esta técnica, un haz luminoso responde a las características eléctricas de señal del habla, y deja una huella sobre un papel sensitivo a la luz.

