

**KEITH JOHNSON (2003): *Acoustic & Auditory Phonetics*, Blackwell, Oxford  
2a edición.**

Keith Johnson entrega la segunda edición de su *Acoustic & Auditory Phonetics* que se diferencia de la anterior por algunas modificaciones menores en el texto y también por un interesante nuevo capítulo sobre percepción.

A pesar de ser un libro orientado a los estudiantes, la primera sensación que produce es que se requiere una base matemática y física más o menos sólida para su cabal comprensión.

El libro en general trata de tres dimensiones de la fonética: acústica, auditiva y perceptiva. En una primera parte nos presenta la teoría y en la segunda parte se dedica a explicar las características de los sonidos en las lenguas desde esos tres puntos de vista.

Cada capítulo termina con ejercicios que permiten efectivamente sacar el máximo provecho al tema presentado. Además, el lector tiene que dar sus propias definiciones de los términos centrales implicados en el capítulo.

El texto contiene mucha información complementaria al asunto que se está tratando. Estos contenidos adicionales se encuentran en cajas de texto grises, para señalar que es información relacionada y que se puede leer con cierta autonomía.

El primer capítulo presenta los conceptos fundamentales relacionados con la fonética acústica. Tras explicar brevemente que el sonido consiste en la percepción, por parte del oído, de las fluctuaciones de presión de aire producidas por el movimiento de un objeto, Johnson señala que la onda acústica constituye una representación de este movimiento a través del tiempo. A continuación, presenta breves definiciones de diferentes tipos de sonidos, exponiendo las características que poseen las ondas acústicas que los representan.

Las ondas sinusoidales resultan metodológicamente útiles porque permiten describir ondas más complejas como el resultado de la suma de varias de estas ondas. Estas ondas pueden describirse en función de tres conceptos básicos: *amplitud*, *frecuencia* y *fase*.

Johnson define las ondas periódicas complejas como aquellas que se componen por al menos dos ondas sinusoidales. Estas ondas presentan patrones que se repiten regularmente. A la frecuencia con la que se repiten estos patrones se la llama *frecuencia fundamental* o  $f_0$ . A continuación, indica que a través de un análisis de Fourier podemos descomponer una onda compleja en las ondas sinusoidales que la constituyen, lo que nos da como resultado un gráfico de frecuencia y amplitud llamado *espectro*.

Por oposición, las ondas aperiódicas no presentan patrones regulares y su espectro tiene una forma totalmente diferente de la de las ondas periódicas. Incluye explicaciones sobre el ruido blanco y los *transient*, o cambios súbitos de presión que no se prolongan en el tiempo.

A continuación, Johnson trata el tema de los *filtros acústicos*. Un filtro permite seleccionar diferentes componentes frecuenciales de una señal. Los conceptos que se presentan aquí son, entre otros, los de centro de frecuencia y ancho de banda. Define y explica los tres tipos de filtro: *low-pass*, *high-pass* y *band-pass*.

En este contexto, resulta muy interesante la definición de frecuencia fundamental como máximo común denominador de un conjunto de armónicos. Efectivamente, cuando uno crea una señal periódica compuesta, el fundamental no es necesariamente la frecuencia más baja, sino el divisor común; por ejemplo, si los componentes son de 500 Hz y 600 Hz, el fundamental será de 100 Hz (aún cuando

---

no lo hayamos incluido expresamente en los componentes). Esta característica de las señales es lo que permite que la línea telefónica transmita en una banda limitada de frecuencias y que, a pesar de ello, siempre se identifique la melodía del habla.

El capítulo 2 está centrado en lo que es el procesamiento digital de la señal y comienza exponiendo la diferencia entre sonido digital y analógico. Se presenta el sonido digital como un tipo de archivo legible por ordenador que se construye a partir de dos simplificaciones realizadas sobre el sonido analógico: *sampling* y *quantization*. La primera de estas simplificaciones se produce en el eje temporal de la señal original y consiste en el número de muestras que se toman por segundo. Cuantas más muestras por segundo se tomen, evidentemente el sonido digital será más parecido al analógico, pero tanto más grande será el archivo. Una cifra adecuada para los propósitos de la investigación fonética es 22050 muestras por segundo.

Dado que para representar una frecuencia, se requieren al menos dos muestras por ciclo, con una determinada frecuencia de muestreo tendremos información frecuencial hasta la mitad de ese valor. Por ejemplo, con 22050 muestras por segundo tendremos información hasta alrededor de los 11000 Hz

La segunda simplificación, o el segundo proceso en la digitalización de la señal analógica, es la llamada *quantization* y que algunos traducen como *formato de muestreo* (por oposición a la *frecuencia de muestreo*). Esta es la simplificación en la amplitud de la señal. Lo relevante aquí es cuántos pasos se distinguen para formalizar las variaciones en la amplitud. También es evidente que mientras más pasos se distinguen, mayor fidelidad tendrá el sonido digital con respecto al analógico. En los ordenadores, estos pasos se representan con el número de bits; con un formato de 16 bits se pueden representar suficientemente bien las variaciones de amplitud.

Dentro de este capítulo se presentan también los conceptos de ventana de análisis, método de autocorrelación para la determinación del *pitch*, análisis LPC, formas de tratamiento de la amplitud de la señal, los filtros desde el punto de vista del tratamiento digital, y las nociones de espectro y de espectrograma.

El tercer capítulo está dedicado a unas notas sobre el mecanismo auditivo. El autor trata tres aspectos: la anatomía y fisiología del oído; la percepción de la intensidad y de la frecuencia; y los modelos auditivos mediante ordenadores.

En cuanto al primer punto, explica la anatomía y la fisiología de lo que se llama sistema auditivo periférico, que se ocupa de transformar señales acústicas en impulsos neuronales. Abarca desde el oído externo hasta el nervio auditivo. Resumidamente, lo que sucede en este sistema es lo siguiente: los impulsos acústicos llegan al tímpano, que es una membrana flexible que reacciona a esos impulsos como la membrana de un tambor. Estos impulsos son transmitidos al oído medio que está formado por una cadena de huesecillos, que, a su vez, transmite los citados impulsos al oído interno. Este está compuesto de la cóclea en cuyo interior se encuentra la membrana basilar. El movimiento de los huesecillos del oído medio llega al oído interno gracias a que este se encuentra lleno de líquido. La membrana basilar tiene diferentes grosores, y la reacción a las diferentes frecuencias depende del grosor de cada parte; por esa razón, en la zona cercana a los huesecillos donde la membrana es más gruesa, reacciona mejor a los sonidos de baja frecuencia. Diferentes fibras del nervio auditivo inervan diferentes zonas de la membrana basilar de manera que al cerebro llega información de las distintas frecuencias a través de estas fibras.

En la segunda parte explica cómo el sistema auditivo periférico trata las señales acústicas. Comienza hablando del tratamiento de la intensidad. Observa que la relación entre presión de sonido y decibelios, por un lado, y la sensación subjetiva del oyente, por otro, no es lineal. La sensación subjetiva suele medirse en sones. La no linealidad se debe a que no siempre el doble de presión de sonido se corresponde con el doble de sones. Esta no linealidad la representan también los decibelios pero hemos de tener en cuenta que, aun con todo, esta unidad es acústica y no perceptiva, de ahí que no coinciden sones y decibelios.

En cuanto al tratamiento de la frecuencia por parte del sistema auditivo periférico, sucede algo bastante similar: una diferencia de 500 Hz, entre 500 y 1000 Hz, es mucho mejor percibida que entre frecuencias de 5000 y 5500 Hz. Este tratamiento diferenciado de las frecuencias del sonido por parte del oído interno se debe justamente a la configuración específica de la membrana basilar, porque esta membrana tiene una mayor superficie que reacciona a frecuencias bajas. Para representar estas frecuencias “auditivas” se usa la escala Bark que es diferente de la de frecuencias “acústicas” (en Hz).

Por último, Johnson se ocupa de los modelos auditivos. Existen diversos modelos de representación psicoacústica con los cuales se pueden realizar los gráficos equivalentes a los espectros y a los espectrogramas (se usa en estos casos la escala de frecuencia expresada en Barks). En todas estas representaciones puede observarse la diferencia entre el mundo acústico y el perceptivo, porque las

---

frecuencias bajas suelen corresponder a una porción muy grande de una escala tradicional establecida en Hz. En la presentación de esta materia, el texto de Johnson presenta convincentes ejemplos en los que se observa que las frecuencias bajas son muchísimo más relevantes desde el punto de vista perceptivo. Estas simulaciones psicoacústicas, equivalentes a los espectros y a los espectrogramas tradicionales, representan bastante bien el modo en que los oyentes identifican y distinguen los sonidos.

El capítulo concluye con los comentarios a un espectrograma y a su equivalente “cocleagrama”.

En el capítulo 4, Johnson establece las diferencias entre representación perceptiva y representación auditiva a partir de dos premisas:

1. Auditivamente, los sonidos del habla son configurados por el sistema auditivo, mientras que perceptivamente no se limitan solo a la audición, sino que incluyen la visión.
2. La percepción de los sonidos del habla se dan en el contexto de la experiencia del hablante-oyente en una lengua determinada.

En este capítulo, presenta una metodología relativamente sencilla para estudiar la percepción del habla, y además muestra que la percepción es el resultado de la audición más otra información.

Para el autor, la investigación acerca de la percepción del habla debe analizar la estructura del “espacio perceptivo” de los oyentes, basándose en las confusiones que estos manifiestan en el nivel auditivo.

Keith Johnson presenta de manera muy didáctica los pasos que se deben seguir para hacer una investigación en esta área. Estos pasos incluyen desde la selección del material que se va a estudiar (pasando por especificaciones de cómo se debe agregar ruido a las señales para realizar los tests), hasta la creación de una matriz de confusión con las repuestas obtenidas por los informantes.

Para extraer información de una matriz de confusión, Johnson utiliza el método llamado *multi-dimensional scaling*, MDS, que es una extensión de la triangulación. Esto permite elaborar un “mapa perceptivo”, el que da cuenta de la estructura del

“espacio perceptivo” de los oyentes. Para elaborar este mapa perceptivo se deben llevar a cabo dos pasos.

1. Calcular la similitud, de acuerdo con la cantidad de veces que el sonido A es confundido con el sonido B y viceversa.
2. Derivar distancias perceptivas a partir de la similitud.

No es este el lugar para dar el detalle de los procedimientos señalados por el autor, pero sí para mencionar que su presentación en este punto es extraordinariamente sistemática y aplicable.

De la presentación de Johnson, se desprende que la confusión perceptiva entre dos sonidos en determinada lengua no tiene que ver solamente con la audición. El mapa de audición se realiza utilizando solamente estímulos auditivos, mientras que el mapa de percepción debe considerar el sonido y la imagen. El llamado efecto McGurk confirma este planteamiento: si se combina el audio de un sonido con la imagen de una persona pronunciando otro (y ambos son parecidos) el observador percibe un tercer sonido.

Basándose en un corpus significativo de investigaciones, Johnson establece que los hablantes de diferentes lenguas poseen diferentes mapas perceptuales. Esto significa que, mientras los patrones auditivos son universales (todos los seres humanos del mundo tenemos oídos muy similares y, por lo tanto, el proceso de percepción del sonido es muy similar también), los patrones perceptivos son particulares de cada lengua.

El capítulo quinto es la exposición de la teoría acústica de la producción del habla. A lo largo del capítulo usa el ejemplo de la vocal neutra como caso prototípico para toda la ejemplificación correspondiente.

El capítulo comienza con la explicación de la fuente sonora y de los componentes armónicos. Continúa estableciendo la analogía con la pulsación de una cuerda de guitarra y sus modos de vibración para establecer firmemente la relación entre el componente fundamental y los armónicos. El autor hace una relación entre la sonoridad y la llamada *quantal theory* ya que en el caso de la actividad de las cuerdas vocales es fácil identificar las zonas de estabilidad (*voiceless*, *voiced* y *glottal stop*).

---

Continúa con la función de filtro del tracto vocal, que cambia la señal constituida por el fundamental y los armónicos (en tanto fuente). Con ello y los modelos de resonancias de tubos, el autor avanza hacia la noción de formante para concluir esta sección del capítulo con las fórmulas de las resonancias de los tubos cerrados en un extremo. Cuando se relacionan estos modelos de resonancias de tubos con el tracto resonador humano se comprende bien la noción de formante. Siguiendo las fórmulas presentadas por el autor para calcular resonancias en un tubo cerrado en un extremo y sabiendo también la longitud del tracto resonador (la longitud es parte de la fórmula), podemos calcular sus resonancias o formantes. La vocal neutra tiene la ventaja de que su descripción articulatoria es muy simple y se puede modelar como si se tratara de un tubo sin obstáculos; de esta manera, un tubo de 17.5 cm tiene las más bajas resonancias a los 500 Hz y a los 1500 Hz

A partir del capítulo seis, el autor comienza con la aplicación de todo lo mencionado anteriormente. El punto de partida son las vocales cuyas resonancias o formantes se explican a partir del modelo de tubos. También Johnson muestra unos nomogramas (en este caso se trata de gráficos que tienen en el eje horizontal el parámetro articulatorio de localización de la constricción y en el eje vertical el parámetro acústico de la frecuencia) para explicar cómo puede cambiar la filiación de la cavidad resonante asociada a determinado formante en la medida en que cambia la posición de la constricción.

Otra manera de explicar las vocales es siguiendo la teoría de la perturbación, que señala que hay puntos de máxima velocidad y puntos de máxima presión en el tracto vocal. Así, si la constricción se produce en un punto o en otro del tracto los efectos sobre las moléculas de aire serán inversos. Las constricciones cerca de un punto de máxima velocidad hace que los formantes tengan valores más bajos; a la inversa, constricciones en puntos de máxima presión elevan la frecuencia de los formantes.

A partir de la teoría de la dispersión adaptativa, de Lindblom, el autor hace una interesante reflexión acerca de la posibilidad de explicar por qué las vocales de las esquinas del espacio en el que se representan típicamente, tienden a ser las preferidas en muchas lenguas. Y la razón aducida es que en esas posiciones se sitúan vocales cuyos formantes son, en relación con las demás vocales, más fácilmente distinguibles. En esta parte del texto el autor desarrolla una muy interesante discusión entre esta teoría de la dispersión adaptativa de Lindblom y la *quantal theory* de Stevens.

Finaliza el capítulo con la perspectiva auditiva y perceptiva de las vocales mostrando, por una parte, espectros auditivos y acústicos superpuestos para constatar las diferencias entre estas dos perspectivas y, por otra parte, estudios de percepción de vocales de los que se derivan mapas perceptuales diferentes para lenguas que tienen, aparentemente, las mismas vocales.

El séptimo capítulo del texto está concentrado en los ruidos fricativos, caracterizados como aperiódicos, en los que la fricción es el resultado de la combinación de dos factores: el tamaño del canal y la velocidad del flujo de las moléculas.

También se aplica el modelo de tubos a las fricativas. La frecuencia más alta de un ruido de esta naturaleza depende del tamaño de la cavidad frontal (desde el punto de constricción hasta los labios): cuanto mayor sea el tamaño de la cavidad, menor frecuencia tendrá el sonido. (Esto no se aplica a los sonidos que no presentan cavidad anterior como los labiodentales que presentan más bien energía distribuida en una amplio gama de Hz.)

El autor hace una comparación entre las representaciones espectrales auditiva y acústica, y constata cómo algunas dificultades que aparecen en el análisis que se hace a partir de la información acústica tienden a desaparecer cuando se toma como punto de partida la información de naturaleza auditiva. Especialmente llamativo resulta el hecho de que las frecuencias sobre los 5 kHz no tienen relevancia perceptiva para el oyente.

Como en todos los casos, el autor reflexiona acerca de la distribución perceptiva de estos ruidos. En este caso confronta el mapa perceptual de las fricativas del japonés y del inglés estadounidense, a partir de datos proporcionados por otras investigaciones.

En el capítulo ocho, Johnson estudia los sonidos oclusivos y africados abarcando las distintas posibilidades que encontramos en las lenguas del mundo. Explica las diferencias con unos pocos parámetros acústicos. En primer lugar, destaca que, a diferencia de las vocales y las fricativas, en las oclusivas el tracto vocal pasa por un movimiento de cierre y otro de liberación. Esta característica resulta clave a la hora de entender las propiedades acústicas de los sonidos oclusivos y africados.

Johnson considera dos aspectos a la hora de estudiar estos sonidos. Por un lado, la fuente del sonido y, por otro, las funciones del tracto vocal como filtro. Con respecto al primero comienza señalando las diferencias de los diversos tipos de

fonación en la producción de estos sonidos. Así, se producen diferencias acústicas perceptibles en los espectros y en los oscilogramas, dependiendo del estado de la glotis: modal, *creaky* o *breathy*. Las diferencias vienen dadas por el cociente de apertura: el tiempo en que la glotis permanece abierta con respecto al tiempo en que se encuentra cerrada. Este cociente aumenta en el caso de la voz *breathy* y disminuye en el caso de la voz *creaky*.

Hablando ya propiamente de las fuentes de sonido, señala Johnson que son diferentes en el momento de cierre y en el de la explosión. Así, en el primer caso la única fuente posible son las cuerdas vocales, mientras que en el segundo encontramos otros tipos de fuente de sonido. Cuando se produce el cierre del tracto vocal, la presión subglotal y supraglotal se igualan con lo que se dificulta cualquier movimiento de las cuerdas vocales. Los resultados con respecto a este problema pueden ser dos: bien solo hay sonorización durante un periodo del tiempo de cierre, o bien se produce alguna maniobra para favorecer el movimiento de las cuerdas vocales. La primera solución es la que se observa en los sonidos propiamente oclusivos sonoros; en sus espectrogramas podemos observar un desvanecimiento paulatino de la barra de sonoridad a lo largo del periodo de silencio. Si se producen maniobras de expansión de la cavidad oral para facilitar el aumento de la presión subglotal, el resultado es un sonido inyectivo.

En cuanto a las fuentes de sonido en el momento de la explosión, tenemos que considerar el momento propio de liberación del aire. Este también puede ser sonorizado del mismo modo que lo era el periodo de cierre. Los dos ruidos son diferentes porque sus puntos de producción son distintos. También son diferentes su duración y los filtros vocales que se aplican a ambos sonidos. Las diferencias acústicas entre sonidos oclusivos, eyectivos y clics radican en la intensidad de la explosión, mayor en los últimos debido a su escaso volumen de aire atrapado en la cavidad bucal.

Las cavidades bucales varían durante la producción de los sonidos oclusivos. En unos momentos, la cavidad anterior y posterior al cierre están emparejadas y, en otros, prevalece una de ellas sobre la otra. Al tratarse de órganos en movimiento y no de posturas estáticas (como en sonidos fricativos), la información sobre el punto de articulación se encuentra en las transiciones de los formantes y no en sus valores medios. Especialmente relevante en este sentido resulta el concepto de *locus* comentado por el autor a partir de los trabajos de Delattre *et al.* A partir del *locus* se puede determinar con relativa facilidad el punto de articulación de estos sonidos. El siguiente punto trata de la diferencia entre un sonido africado y una combinación de un sonido oclusivo seguido de uno fricativo. La distinción acústica

se encuentra en lo que se ha llamado tiempo de elevación (*rise time*). La amplitud del ruido de fricción crece mucho más deprisa en el caso de los sonidos africados.

Finalmente, el autor se ocupa de asuntos auditivos y perceptivos relacionados con estos sonidos. Auditivamente, los cocleagramas presentan dos diferencias importantes con respecto a los espectrogramas. En primer lugar, en los cocleagramas se observan resaltadas las transiciones del segundo formante. Además, las explosiones son más prominentes en los cocleagramas de lo que lo son en los espectrogramas. En cuanto a la percepción de estos sonidos comenta una interesante hipótesis sobre la generalizada palatalización del sonido velar [k]. En ella, no solo se reconoce la importancia de la articulación sino que también es bastante probable que la percepción desempeñe un papel crítico.

El capítulo nueve está dedicado a las nasales y a las laterales. Dadas las cualidades de estos sonidos, el autor comienza presentando la noción de ancho de banda y explicando su relación con las cavidades resonantes; también se refiere al fenómeno de la antirresonancia (y a los llamados antiformantes) y lo explica en términos de sus correlatos articulatorios. También presenta una posible interpretación en términos perceptuales de las nasales.

Para concluir esta reseña, resulta necesario destacar algunas ideas y presentarlas en forma sumaria:

El libro de Johnson está centrado en la fonética segmental y, en principio, está destinado a estudiantes; no obstante requiere un lector con una buena base en física y matemática.

El libro tiene un énfasis especial en el sonido digital y en el enfoque perceptivo. Presenta primero la teoría y, en una segunda parte, las cuestiones más aplicadas. En todo momento, el libro mantiene la coherencia dada por el constante contrapunto entre la información acústica y la perceptual.

Los autores de esta reseña consideramos que se trata de un libro en varios aspectos ejemplar y motivador. Para explicar nociones de fonética acústica después de todo el abundante material que existe, hay que tener alguna especificidad en el tratamiento. En este sentido, el libro que reseñamos es de las mejores selecciones de tópicos teóricos y estos son tratados sin eludir problemas. Tiene, además, muchos méritos en el plano didáctico. Por esta razón, su lectura es interesante para todos los que se dedican a la docencia en esta área.

Sin duda, lo más destacable es la demostración persistente de que las visiones acústicas, auditivas y perceptivas se deben considerar complementarias. De esa manera se puede evitar la aproximación tecnicista y se tiende a una fonética más integradora.

*Valeria Cofré Vergara*

Pontificia Universidad Católica de Chile  
vacofre@uc.cl

*Domingo Román Montes de Oca*

Pontificia Universidad Católica de Chile  
dromanm@uc.cl

*Francisco Javier Simón Casas*

Universidad de Zaragoza  
jasimon@unizar.es