



La sílaba: Unidad temporal común en percepción y producción de habla

Elvira Pérez

National Biomedical Research Unit in Hearing, Nottingham, Reino Unido

Tipo de artículo: Clásicos.

Disciplinas: Psicología, Lingüística.

Etiquetas: lenguaje, percepción, producción, sílaba.

La sílaba no es sólo una unidad lingüística en producción de habla, sino también una unidad temporal que organiza nuestra percepción auditiva. Recientes modelos teóricos como la Gramática Auditiva (Nakajima & Sasaki, 1996) resaltan esta idea proponiendo sistemas de producción y percepción independientes con acceso a un mecanismo central común dedicado al procesamiento y organización temporal de ambos sistemas.



(cc) Daniel Montesinos

La sílaba es una unidad lingüística, una división válida que podemos hacer tanto en el habla como en la lectura (véase Domínguez y de Vega, 2008, <http://www.cienciacognitiva.org/?p=54>). Pero la sílaba es mucho más que eso. Es, además, una unidad temporal que organiza y limita tanto la producción como la percepción del habla y otros sonidos. En producción, la sílaba tiene una duración media de unos 160-300 ms, independientemente de la lengua o idioma que se analice. En percepción sucede algo similar, los sonidos se organizan en ventanas temporales de unos 160-300 ms. En estas ventanas temporales los sonidos se descomponen para más tarde ser organizados siguiendo las reglas impuestas por el sistema auditivo. A veces, estas reglas alteran el orden lineal en el que se originaron los sonidos creando interesantes ilusiones auditivas. El estudio sistemático de estas ilusiones es una de las herramientas más valiosas que tenemos para

comprender la relación temporal que existe entre la percepción y producción de sonido y habla.

La sílaba en producción

Tradicionalmente, las lenguas se han clasificado según las regularidades temporales con que se producen sus unidades silábicas. Según esta clasificación nos encontramos con lenguas como el español y el japonés, de claro ritmo silábico (“syllable-timed”), o el inglés, de claro ritmo acentual (“stress-timed”). Las lenguas de ritmo silábico se caracterizarían por la alta regularidad temporal de sus unidades silábicas. En cambio, en las lenguas de ritmo acentual la duración temporal de las unidades silábicas varía más.

Esta clasificación, sin embargo, es un tanto simple, y la dicotomía ritmo silábico/acental esconde más similitudes que diferencias. Además, como la regularidad temporal absoluta no existe en producción, algunos autores entienden que ésta es un producto más del plano perceptual. Esto se puede demostrar si analizamos y comparamos las propiedades temporales de dos lenguas de distinto ritmo, como son el japonés y el inglés. Por ejemplo, Arai y Greenberg (1997) transcribieron y segmentaron un gran corpus de habla espontánea en ambas lenguas y encontraron que la duración media de la sílaba japonesa (166 ms) no se diferenciaba de manera significativa de la duración media de la sílaba inglesa (190 ms). Los coeficientes de variación, dispersión y las distribuciones temporales de estas sílabas también mostraban más similitudes que diferencias. Tales similitudes temporales resultan sorprendentes si tenemos en cuenta las grandes diferencias fonológicas de ambas lenguas. Estos patrones temporales parecen reflejar límites universales impuestos por los sistemas articulatorio y perceptual, comunes a todas las lenguas, y nos dan una pista para entender la facilidad con la que el sistema auditivo humano decodifica habla cuando las condiciones acústicas son pobres.

La sílaba en percepción

La sílaba es la unidad lingüística más fácil de identificar, más incluso que los fonemas (Segui, Dupoux, y Mehler, 1990) y hasta en lenguas poco o nada familiares. Una de las teorías más aceptadas para explicar esta facilitación silábica propone que el sistema auditivo está especializado en capturar y organizar la información acústica contenida en intervalos temporales del rango 160-300 ms. Esta ventana temporal facilita la representación estable del habla, ya que refleja los cambios acústicos asociados a los movimientos articulatorios de labios, lengua y mandíbula. Este rango temporal es el idóneo para extraer la información silábica y fonética necesaria para acceder a la representación lingüística inherente en el habla. De esta manera, las propiedades temporales que limitan los mecanismos de la producción (sistema articulatorio) y percepción (sistema auditivo) coincidirían en la sílaba como unidad temporal unificadora.

La sílaba: enlace temporal en percepción y producción

Tanto el sistema articulatorio como el auditivo parecen estar diseñados para producir y detectar sílabas. Sin embargo, los modelos teóricos que combinan percepción y producción de habla son escasos y todavía no está claro si el sistema encargado del procesamiento fonológico tiene dos vertientes independientes para la producción y percepción de habla (Levelt, Roelofs y Meyer, 1999), o una vertiente común (Coleman, 1998). Cada vez son más los estudios neurológicos (Coleman, 1998; Hickok y Poeppel, 2000) que apuntan a la región cortical del lóbulo temporal superior izquierdo como área común en el procesamiento de aspectos fonológicos, tanto en producción como en percepción.

La Gramática Auditiva (Nakajima y Sasaki, 1996) propone la sílaba como unidad temporal común en percepción y producción en un modelo de percepción general que incluye dos pasos. El primero se encarga de procesar sonidos de 30 ± 10 ms de duración relacionados con la detección de elementos básicos de la señal, como son comienzos, terminaciones y silencios. El segundo paso se encarga de organizar estos elementos en ventanas temporales de unos 160-300 ms siguiendo una serie de reglas organizativas impuestas por el sistema auditivo, necesarias para la representación perceptiva de la señal. Este conjunto de reglas lógicas puede producir ilusiones auditivas que, en definitiva, son simplemente correcciones. Por

ejemplo, una de las reglas es que todo comienzo (en inglés 'onset') tiene un final ('offset'). Si dos comienzos o clicks de 30 ms de duración se presentan muy cerca el uno del otro, en un intervalo menor de 160 ms, el sistema auditivo los fusionará y, como resultado, sólo se percibirá un click. Si el intervalo entre clicks es mayor de 160 ms, el sistema auditivo insertará los correspondientes finales y el silencio entre clicks, y el resultado perceptivo serán dos clicks. El objetivo final del sistema auditivo es darle sentido a la información acústica que recibe en ventanas temporales que poseen la misma duración que la sílaba en producción.

Esta conexión intrínseca entre producción y percepción de habla podría apuntar a que la evolución del habla depende no sólo de los límites impuestos por el sistema fonador, sino también de los límites impuestos por el sistema auditivo. En definitiva, hablar y escuchar es más parecido de lo que se cree.

Referencias

Arai, T., & Greenberg, S. (1997). The temporal properties of spoken Japanese are similar to those of English. *Eurospeech*, 2, 1011-1014.

Coleman, J. (1998). Cognitive reality and the phonological lexicon: A review. *Journal of Neurolinguistics*, 11, 295-320.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131-138.

Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, 22, 1-75.

Nakajima, Y., & Sasaki, T. (1996). A simple grammar of auditory stream formation. *Journal of the Acoustical Society of America*, 100, 2681.

Segui, J. Dupoux, E., & Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification, and lexical access. En Altmann, G. (Ed), *Cognitive Models of Speech Processing: Psycholinguistic and computational perspectives*, (pp. 263-280). Cambridge: MIT Press.