

Taula (UIB) núm. 15, 1992

El problema de la modularidad en el procesamiento auditivo y en el procesamiento del habla

Luis Enrique López Bascuas*

Universidad Complutense de Madrid

Resumen

En este trabajo se comentan algunas aportaciones experimentales que han resultado ser relevantes para dilucidar hasta qué punto el procesamiento temprano del habla puede ser considerado un módulo. No se pretende ser exhaustivo a la hora de presentar la evidencia disponible, sino desarrollar una línea argumental coherente que permita establecer algunas conclusiones que ayuden, al menos, a centrar la polémica.

Abstract

In this work relevant experiments are commented concerning the existence of a speech module. The paper is not intended to provide an exhaustive review of the evidence. Instead, it tries to develop a coherent line of argument that may lead to some relevant conclusions in order to set the debate in proper terms.

* Este trabajo ha sido financiado, parcialmente, por la Dirección General de Investigación Científica y Técnica, proyecto PB87-0531. Quiero agradecer a R.P. Fahey y B.S. Rosner (Universidad de Oxford) y a J.E. García-Albea (Universidad Complutense) toda la ayuda prestada en el curso del mismo.

Evidencia primitiva a favor de la existencia de un módulo de procesamiento de habla

El debate en torno al problema de la modularidad (Fodor, 1983) ha tenido una larga tradición en el campo del procesamiento del habla. Ya en los años 50, la singularidad del habla fue defendida enfáticamente en los laboratorios Haskins. Sin embargo no es menos cierto que precisamente en este dominio la hipótesis de la modularidad ha sido duramente criticada.

La evidencia psicológica primitiva que apoyaba la idea de que el habla era «especial» provino de estudios realizados con habla sintética en los ya mencionados laboratorios Haskins. En 1957, Liberman, Harris, Hoffman y Griffith sintetizaron un conjunto de estímulos iguales en todas sus características físicas excepto en la frecuencia de inicio de la transición del segundo formante. Variando en pasos iguales ese valor frecuencial, construyeron un continuo físico que, desde el punto de vista perceptivo, cubría los diferentes lugares de articulación de las oclusivas sonoras (/b/ /d/ /g/). Las frecuencias centrales de las partes estacionarias de los formantes aproximaban la vocal final al sonido /e/. La tarea de los sujetos consistía en identificar los estímulos. Sorprendentemente, aunque los estímulos formaban un continuo físico, los sujetos los clasificaban conforme a tres categorías bien definidas (esto es, las probabilidades de identificación variaban bruscamente en vez de gradualmente). Más aun, Liberman *et al.* trataron de determinar si los sujetos eran capaces de detectar diferencias entre estímulos que pertenecían a la misma categoría. Para ello utilizaron una tarea de discriminación ABX. A y B son dos estímulos diferentes del continuo y X puede ser o A o B. La tarea del sujeto es decidir si X es A o B. Los resultados mostraron que los sujetos sólo podían discriminar entre estímulos que habían sido extraídos de distintas categorías percibidas pero no podían discriminar (al menos, difícilmente podían) entre estímulos que pertenecían a la misma categoría percibida. Este patrón básico de resultados es lo que se conoce como «Percepción Categórica».

Estos hallazgos concernientes al lugar de articulación fueron extendidos a otros contrastes fonológicos. Entre otros, el contraste sordo/sonoro ha sido estudiado extensamente. Existen muchas claves acústicas que potencialmente pueden marcar la distinción sorda/sonora pero una de las más usadas en la investigación sobre habla es el VOT (voice onset time). El VOT es el intervalo que transcurre entre la relajación de los articuladores tras el periodo de oclusión y el inicio de la vibración glotal. Ambos eventos pueden ser localizados en una representación espectrográfica de la señal del habla. La relajación de los articuladores está marcada por un estallido breve de energía en la región de frecuencias medias o altas, en tanto que la actividad glotal se manifiesta en el inicio de la periodicidad en bajas frecuencias. Lisker y Abramson (1970) obtuvieron para esta dimensión el patrón clásico de Percepción Categórica.

Dado que para otro tipo de estímulos los sujetos suelen mostrar un rendimiento mucho mayor en tareas de discriminación que en tareas de identificación y que, para ciertas señales no lingüísticas (por ejemplo Liberman *et al.*, 1961), se encontró percepción continua, la percepción categórica se tomó como evidencia de que algún tipo de procesador especial computaba las señales de habla. Aquí yacen los orígenes de la Teoría Motora de la Percepción del Habla, teoría que en su formulación más reciente (Liberman y Mattingly, 1985) apoya una organización modular de los sistemas de entrada.

Evidencia en contra de un módulo específico: los trabajos con continuos de no habla

Sin embargo, Miller *et al.* (1976) trataron de demostrar que la percepción categórica podría ser una propiedad general de la conducta sensorial, poniendo así en crisis la afirmación de que existe un modo de percepción único para el habla. Para ello construyeron un continuo de no habla con ruido termal filtrado y un zumbido (una función de onda cuadrada) variando en pasos iguales el inicio relativo del ruido y el zumbido. En los experimentos de identificación se presentaron a los sujetos sonidos aislados con asincronías de -10 a +80 ms. Los sujetos tenían que utilizar las etiquetas «no-ruido» o «ruido» para identificar los estímulos. En la tarea de discriminación se presentaba a los sujetos dos estímulos iguales y uno distinto. El sujeto debía indicar si el sonido diferente era el primero, el segundo o el tercero. Los resultados mostraron funciones de identificación con pendientes muy pronunciadas y funciones de discriminación con máximos locales en torno a la frontera entre categorías (en tanto que ejecución cercana al azar en el resto de los puntos).

Pisoni (1977) siguió la misma línea argumental en una serie de cuatro importantes experimentos. Su continuo de no habla trataba de simular otra clave acústica que marca el rasgo de sonoridad, a saber, la frecuencia de inicio del primer formante relativa a la frecuencia de inicio del segundo. Para ello utilizó dos funciones sinusoidales con frecuencias de 500 y 1500 Hz. El continuo se construía variando, en pasos de 10 ms., el inicio del tono grave en relación con el tono agudo. Los puntos extremos del continuo consistían en el tono grave precediendo al agudo 50 ms. y en el tono agudo precediendo al grave 50 ms.

La primera parte del experimento 1 consistió en dos sesiones de entrenamiento donde los sujetos asociaban a cada sonido un botón de respuesta. Acto seguido los sujetos se sometían a una tarea de identificación, usando las respuestas que habían aprendido en la sesión de entrenamiento, y a una tarea ABX de discriminación.

Pisoni obtuvo funciones de identificación como las pronosticadas si se asume percepción categórica al menos para 5 de sus 8 sujetos. Hay que resaltar que la frontera entre categorías no ocurría a los 0 ms. de asincronía sino que estaba desplazada hacia la derecha, esto es, hacia la categoría en la que el tono grave se retarda con relación al agudo. En este punto existen dos posibles explicaciones. Una en términos de enmascaramiento (mayor enmascaramiento de las altas frecuencias por las bajas frecuencias -la amplitud del tono grave era 12 dB mayor que la del agudo-) o una limitación en el procesamiento del orden de eventos temporales. Un experimento piloto no mostró cambios en la frontera entre categorías cuando se variaba la amplitud relativa de los tonos, sugiriendo que la hipótesis en términos de enmascaramiento podía ser descartada.

En las funciones de discriminación aparecieron los picos habituales; así pues se encontró un nuevo ejemplo de percepción categórica con señales de no habla. Sin embargo la cuestión que se suscita es si ese patrón de resultados podría ser el resultado de algún proceso de etiquetado adquirido en las sesiones de entrenamiento. Este es un punto importante porque, de acuerdo con el argumento, los picos deben tener un fundamento puramente psicofísico. Por ello en el experimento 2 los sujetos realizaron la tarea de discriminación ABX sin sesiones de entrenamiento previas.

Los resultados muestran el pico ya encontrado en torno a los +20 ms. Esto significa que los picos no aparecen debido exclusivamente a los procedimientos de entrenamiento, sino que existen áreas en el continuo psicofísico para las cuales el sistema es más sensible. Los datos mostraron también un pico secundario en torno a los -20 ms. Parecería haber tres regiones diferentes: el tono agudo precede al grave 20 ms. o más, el tono grave precede al agudo 20 ms. o más y una región de simultaneidad entre -20 y +20 ms.

Si esto es así es posible avanzar dos predicciones adicionales. En primer lugar, si ofrecemos a los sujetos la posibilidad de utilizar tres etiquetas en vez de dos, deberíamos obtener tres categorías con las fronteras localizadas en -20 y +20 ms. En segundo lugar, si la tarea de los sujetos es decidir si al inicio oyen uno o dos eventos, deberán contestar «dos» sólo cuando los estímulos estén fuera del rango (-20,+20). Ambas predicciones fueron comprobadas en los experimentos 3 y 4.

Todos estos resultados sugieren que la distinción sorda/sonora podría estar basada en las capacidades generales del sistema auditivo, desestimando así cualquier explicación basada en módulos específicos de procesamiento del habla.

Así, Miller *et al.* propusieron una explicación del fenómeno de la percepción categórica en términos de un umbral de enmascaramiento. Si se toman dos estímulos que están por debajo del umbral, no podrán ser distinguidos y los sujetos contestarán al azar, si ambos estímulos están por encima del umbral, entonces podrán ser detectados y las discriminaciones de los sujetos seguirán la ley de Weber. El pico en la función discriminación aparece cuando se toma un estímulo que está por debajo del umbral y otro que está por encima. Esto marcaría una discontinuidad natural que haría sencillo el establecer etiquetas (Cf. Miller *et al.*, 1976).

Por su parte, Pisoni sugirió que el fenómeno de la percepción categórica refleja una limitación básica del sistema auditivo para la estimación del orden en que se producen eventos temporales. Habría una suerte de umbral para este tipo de tareas en torno a los 20 ms. Esto es consistente con los hallazgos de Hirsh (1959) (aunque véase Rosen y Howell, 1987). Los distintos sistemas fonológicos, por tanto, habrían aprovechado esta situación para situar las categorías fonológicas en distintos lados de la discontinuidad natural, con lo que las harían claramente discriminables (Cf. Pisoni, 1977).

Con independencia de los detalles, todas estas explicaciones de la distinción sorda/sonora apoyan fuertemente la idea de una facultad horizontal encargada de tratar todo tipo de señales acústicas.

Nueva evidencia a favor de un módulo específico de procesamiento del habla

Sin embargo, existe también un buen cúmulo de información en el que no encaja la asunción de que barreras psicoacústicas naturales explican las fronteras de VOT. Para empezar por lo más obvio hay que señalar que tales fronteras no son fijas sino altamente flexibles. Varían con el contexto que proporcionaría otro estímulo presentado con el estímulo test (efectos de contraste, de rango, de adaptación selectiva) y también varían con las peculiaridades de la estructura interna del estímulo (efectos de la velocidad de habla y relaciones de intercambio entre claves acústicas) (Repp y

Liberman, 1987). Por ejemplo, es sabido que se da una relación de intercambio entre VOT y la frecuencia de inicio del primer formante. Así la frontera de VOT se alarga según se disminuye la frecuencia de inicio del primer formante. Este resultado no se ha obtenido con análogos de no habla (Summerfield, 1982; aunque véase Diehl, 1987).

Otra fuente de evidencia importante que permite dilucidar si el efecto de la frontera fonémica se debe a una facultad horizontal o vertical proviene del campo de la investigación translingüística. Parece claro que, si algún tipo de umbral psicoacústico está involucrado en el efecto, entonces la posible localización de la frontera fonémica estaría altamente constreñida para los distintos idiomas. Así, un primer paso en la investigación consistió en determinar dónde de hecho sitúan las fronteras los hablantes de diversas lenguas para los distintos contrastes fonológicos. Las diferencias entre el inglés y el español son particularmente interesantes para el problema que nos ocupa. En contraste con su contrapartida inglesa, la oclusiva sonora española se produce con actividad glotal antes del periodo de oclusión de los articuladores, esto es, la /b/ española es presonora. Williams (1977), estudió tanto los aspectos de producción como los de percepción del contraste sordo/sonoro en español (dialectos latinoamericanos) y encontró que, primero, se daba una estrecha correspondencia entre los aspectos de percepción y producción y, segundo, que las fronteras entre categorías y los picos de discriminación diferían de los que se habían descrito en hablantes ingleses.

En particular los ingleses exhibían las fronteras y picos en torno a un VOT de 20 ms. Sin embargo, para el grupo de español ambos valores estaban en torno a -5 ms. Estos resultados no son consistentes con los argumentos presentados anteriormente en tanto que, en ese rango de valores de VOT, se espera una ejecución en discriminación cercana al azar. No obstante, los hablantes de español mostraron un pico secundario en torno a los 20 ms. Parece, por tanto, que pudieron discriminar entre oclusivas aspiradas y no aspiradas en términos puramente auditivos, en tanto que la discriminación correspondiente a los -5 ms. de VOT podría ser de naturaleza fonémica.

Un avance importante cara a esclarecer estas cuestiones se puede encontrar en un trabajo reciente de Kewley-Port, Watson y Foyle, 1988. Estos autores se propusieron mostrar que el fenómeno de la percepción categórica y las capacidades últimas de resolución del sistema auditivo son aspectos independientes en buena medida. Esto es así tanto para señales de habla como para señales de no habla.

Según estos y otros autores conviene distinguir entre dos tipos de tareas: tareas de alta incertidumbre y tareas de mínima incertidumbre psicofísica. Las primeras suponen la presentación de una gran variedad de estímulos en orden aleatorio, en las segundas un mismo par es presentado dentro de una serie de ensayos. En tanto que las primeras serían recomendables cuando estamos interesados en el modo habitual de percibir sonidos complejos, las segundas serían más convenientes si nuestro objetivo es determinar las capacidades básicas (el poder de resolución) del sistema auditivo. Sachs y Grant (1976) ya habían mostrado cómo aumenta el rendimiento usando tareas de mínima incertidumbre en un continuo de VOT con consonantes velares.

Así pues, el argumento es bastante directo: si la ejecución de los sujetos cuando detectan cambios en VOT sigue la ley de Weber (esto es, necesitamos un incremento mayor de estimulación para notar un cambio según aumentamos en VOT) entonces la

percepción categórica no puede explicarse en términos puramente psicoacústicos (por ejemplo, aludiendo a las capacidades de resolución temporal del sistema auditivo).

Cuando usaron las tareas habituales de identificación y ABX obtuvieron los resultados convencionales. Sin embargo cuando usaron una tarea de incertidumbre mínima encontraron una mejora general en las discriminaciones y, lo que es más interesante, una discriminación cada vez más fina según se acortaba la duración del VOT. Este resultado es consistente con la ley de Weber.

Siguiendo los mismos procedimientos repitieron el experimento con un continuo de ruido-zumbido similar al usado por Miller *et al.* (1976). Una vez más se obtuvieron amplias mejoras en ejecución y las funciones eran monótonas decrecientes.

Estos resultados muestran claramente que no hay evidencia fuerte que permita sostener la idea acerca de un umbral psicoacústico que limita la discriminación del orden temporal y, por tanto, otras explicaciones han de ofrecerse para dar cuenta del fenómeno de frontera entre categorías.

Algunas consideraciones en torno al módulo auditivo

Sin embargo, ahora surge una pregunta inesperada: ¿qué determina las fronteras y los picos que se obtienen con sujetos no entrenados en continuos de no habla? De acuerdo con los resultados comentados de Kewley-Port *et al.*, las fronteras podrían haber aparecido en un rango amplio de valores. Dado que de hecho han aparecido en posiciones semejantes a las que se obtienen con habla, una posible hipótesis es que la experiencia con el idioma determina las fronteras y picos en el no habla. Pero si esto es así, información específica del módulo lingüístico está siendo utilizada por el módulo auditivo con lo cual la supuesta modularidad queda violada asumiendo que las computaciones que originan las fronteras son internas a los módulos.

Si tenemos en cuenta los resultados de los estudios translingüísticos citados, tenemos un modo bastante directo de comprobar si la hipótesis de que la experiencia con el idioma determina las fronteras del no habla es sostenible. Como dijimos, las fronteras entre categorías para hablantes de español se sitúan en un VOT menor que las de los hablantes de inglés. Así pues, si la hipótesis es cierta hay que esperar que las fronteras en continuos de no habla sean también menores en hablantes españoles. Fahey, López-Bascuas, García-Albea y Rosner (en preparación) pusieron a prueba esta hipótesis. Aunque el trabajo no está completado, merece la pena comentar los resultados provisionales obtenidos.

En este trabajo se utilizaron dos grupos de sujetos (monolingües españoles y monolingües ingleses) y tres continuos estímulares (uno de habla y dos de no habla).

El continuo de habla estaba constituido por 10 estímulos que perceptivamente variaban de /ba/ a /pa/. El continuo fue generado con una versión modificada del sintetizador de Klatt (1980) desarrollada en el I.B.M. UK Science Center. El punto extremo correspondiente al estímulo /ba/ tenía 35 ms. de pre sonoridad, el punto extremo correspondiente a /pa/ tenía un VOT de 55 ms. Las transiciones de los tres primeros formantes no eran lineales con el objeto de obtener habla más natural. F4 y F5 se mantuvieron constantes.

Los continuos de no habla trataron de simular los de ruido-zumbido y tonos ya comentados. El continuo de ruido-zumbido fue también creado con el sintetizador. Para obtener el ruido se excitó un filtro paso-todo con ruido de fricción; el zumbido se consiguió con una onda cuadrada de 100 hz. En un extremo del continuo el zumbido precedía al ruido en 35 ms., en el otro extremo el ruido precedía al zumbido en 55 ms.

Los tonos fueron creados con un programa en Pascal. El programa permitía definir la asincronía entre dos funciones sinusoidales, una de 500 Hz. y la otra de 1500 Hz. De nuevo los puntos extremos eran de 35 ms. (cuando el tono grave precedía al agudo) y 55 ms. (cuando el tono agudo precedía la grave).

Los sujetos pasaron por 200 ensayos de identificación y 360 ensayos de discriminación bajo el paradigma 2IAX (se presentaban dos estímulos del continuo y el sujeto tenía que decidir si eran iguales o diferentes). Sin entrar en detalles, los resultados hasta ahora disponibles indican que: 1) Existen diferencias significativas entre las fronteras de habla del grupo de españoles y el grupo de ingleses, por un lado, y entre el continuo de ruido-zumbido y el continuo de los tonos en el grupo de españoles, por otro lado. El resto de las comparaciones posibles arrojan diferencias que no son significativas. 2) Ninguna de las seis posibles correlaciones entre continuos (tres dentro de cada grupo de sujetos) resulta ser significativa.

La discusión de estos resultados conviene integrarla en el siguiente epígrafe de consideraciones generales.

Consideraciones finales

El trabajo anterior trataba de decidir si era plausible que la experiencia con un idioma determinara las fronteras que se establecen en continuos de no habla. Si así fuera no debería haber diferencias entre fronteras pertenecientes a distintos continuos dentro de cada idioma. Como hemos señalado el trabajo citado pone de manifiesto un contraejemplo a esta condición: los tonos y el ruido-zumbido difieren significativamente en el grupo de español.

Además, teniendo en cuenta que las fronteras de habla difieren significativamente, habría que encontrar diferencias también significativas entre los continuos de no habla a través de los idiomas. De nuevo encontramos un contraejemplo de esta condición: las fronteras en el ruido-zumbido no son diferentes a través de los idiomas.

Así pues, por un lado parece implausible que el efecto de frontera entre categorías pueda ser debido a propiedades generales del sistema auditivo y, por otro lado, no se ha encontrado evidencia convincente que apoye la idea de que las fronteras de no habla vienen determinadas por la experiencia con el lenguaje. Por lo tanto, la evidencia que aquí se presenta es compatible con una organización modular de los sistemas de entrada.

No obstante, sería poco sensato obviar la evidencia que apunta en la otra dirección así como algunas cuestiones que quedan pendientes de respuesta. Al menos dos pasos deben darse para tratar de resolver el conflicto que plantea la evidencia disponible.

En primer lugar, habría que retomar con rigor la idea de que en los resultados experimentales se mezclan tanto factores sensoriales como factores decisionales. Es

este un aspecto central que no puede obviarse al interpretar los datos. En este sentido nótese que el efecto de frontera entre categorías podría deberse a cualquiera de los dos factores aludidos. Por ejemplo, podría ocurrir que las densidades de la señal que corresponden a las categorías del habla estuvieran claramente separadas o que dichas densidades tuvieran una varianza pequeña. No obstante, también podría ocurrir que las categorías de habla generaran un criterio de respuesta de varianza pequeña en la frontera (Rosner, en prensa).

Los métodos convencionales que siguen el modelo de la teoría de detección de señales tienen una seria limitación: asumen que la varianza del criterio es uniforme y habitualmente cero. Por tanto, es necesario un nuevo modelo que permita varianzas no uniformes del criterio de respuesta. Recientemente se ha hecho un avance importante en esa dirección (Rosner, en prensa). Este tipo de análisis podría revelar diferencias en el procesamiento de señales de habla y no habla.

En segundo lugar, habría que afrontar el problema de encontrar una representación temprana adecuada para las señales de habla y de no habla. Este paso es esencial para desarrollar modelos fuertes de procesamiento auditivo y lingüístico. Los procesos que hayan de ser ejecutados en estadios ulteriores de procesamiento podrían depender críticamente del formato y la información especificados en esa primera representación del estímulo físico. Además, algunos problemas importantes como el de la variabilidad de la señal, podrían resolverse, al menos parcialmente, si se desarrollaran algoritmos que proporcionaran la representación requerida. Más aún, diferentes sistemas de entrada podrían ser caracterizados por los diferentes tipos de representaciones que computan con el fin de resolver el problema perceptivo con el que se enfrentan. Por supuesto, este punto es especialmente relevante para el caso de las señales acústicas y de habla, en tanto que potencialmente podrían compartir muchas características físicas. Así, para identificar a qué dominio pertenece una señal, se debería examinar el tipo de representación inicial que computa.

La naturaleza específica de la representación tiene que determinarse, por un lado, haciendo referencia a la tarea computacional que el sistema ejecuta (esto es, por referencia al problema que el organismo resuelve) y, por otro lado, refiriéndose al conjunto de constricciones que operan sobre los procesos posibles (Marr, 1982). Así existen constricciones matemáticas (cierto teorema podría avisarnos de que una determinada transformación formal provoca un resultado no deseado), constricciones físicas (relativas a la naturaleza del estímulo que ha de ser procesado) y constricciones informativas (qué información puede consultar el proceso).

Por ejemplo, es bien sabido que las claves acústicas que marcan contrastes relevantes en el habla están distribuidos tanto en el dominio del tiempo como en el de la frecuencia. Por tanto, cualquier representación del habla, para ser adecuada, ha de ser máximamente eficiente en ambos dominios. Si se prueba que el procesamiento auditivo general requiere un tipo distinto de representación, tendremos un argumento fuerte para postular dos módulos diferentes.

Bibliografia

- DIEHL, R.L. (1987). Auditory Constraints on Speech Perception. En M.E.H. SCHOUTEN (Ed.), *The Psychophysics of Speech Perception*. Martinus Nijhoff.
- FAHEY, R.P., LOPEZ-BASCUAS, L.E., GARCIA-ALBEA, J.E. & ROSNER, B.S. (en preparación). Temporal lag perception in Spanish and English speakers.
- FODOR, J.A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- HIRSH, I. (1959). Auditory perception of temporal order. *JASA*, 31, 759-766.
- KEWLEY-PORT, D., WATSON, CH. & FOYLED. (1988). Auditory temporal acuity in relation to category boundaries: speech and nonspeech stimuli. *JASA*, 83 (3), 1133-1145.
- KLATT, D. (1980). Software for a cascade-parallel formant synthesizer. *JASA*, 67, 971-995.
- LIBERMAN, A.M., HARRIS, K.S., HOFFMAN, H.S. & GRIFFITH, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.
- LIBERMAN, A.M., HARRIS, K.S., KINNEY, J.A. & LANE, H. (1961). The discrimination of relative onset times of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 61, 379-388.
- LIBERMAN, A.M. & MATTINGLY, I.G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- LISKER, L. & ABRAMSON, A. (1970). The voicing dimension: some experiments in comparative phonetics. Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 563-567.
- MARR, D. (1982). *Vision*. Freeman.
- MILLER, J.D., WIER, C., PASTORE, R. KELLY, W. y DOOLING, R. (1976). Discrimination and labeling of noise-buzz sequences with varying NLT: an example of categorical perception. *JASA*, 60, 410-417.
- PISONI, D.B. (1977). Identification and discrimination of the relative onset times of two component tones: implications for voicing perception in stops. *JASA*, 61, 1352-1361.
- REPP, B.H. & LIBERMAN, A.M. (1987). Phonetic category boundaries are flexible. En S. HARNAD (Ed.), *Categorical Perception*. Cambridge, U.P.
- ROSEN, S. & HOWELL, P. (1987). Is there a natural sensitivity at 20 ms. in relative tone-onset-time continua? A reanalysis of Hirsh's (1959) data. En M.E.H. SCHOUTEN (Ed.) *The Psychophysics of speech perception*. Martinus Nijhoff.
- ROSNER, B.S. (en preparación). Thurstonian models for changes in perceptual performance.
- SACHS, R.M. & GRANT, K.W. (1976). Stimulus correlates in the perception of voice onset time (VOT): II Discrimination of speech with high and low stimulus uncertainty. *JASA*. Supplement 1, 60, S91.
- SUMMERFIELD, Q. (1982). Differences between spectral dependencies in auditory and phonetic temporal processing: relevance to the perception of voicing in initial stops. *JASA*, 72, 51-61.
- WILLIAMS, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics*, 5, 169-184.