

La percepción del habla: problemas y restricciones computacionales

Luis E. López-Bascuas
Universidad Complutense de Madrid

Este trabajo tiene un doble objetivo. Por un lado pretendemos establecer claramente cuáles son los problemas inherentes a la percepción del habla, por otro, se comentarán las posibles restricciones informativas de las que podría servirse un procesador de habla para solucionar eficientemente dichos problemas. Los problemas principales de la percepción del habla se derivan de la relación compleja que existe entre la señal acústica y el mensaje fonológico que se recupera a partir de ella. Las restricciones informativas podrían venir de dos ángulos diferentes: de las constricciones psicofísicas impuestas por el funcionamiento del sistema auditivo o de constricciones lingüísticas específicas que actuarían sólo ante señales de voz. También podría ocurrir que la percepción del habla dependa de transformaciones auditivas que son específicas para el tratamiento de señales de voz. Es una cuestión empírica determinar qué conjunto de restricciones son las utilizadas por los seres humanos cuando se enfrentan a la tarea de percibir habla.

Palabras clave: Percepción del habla, variabilidad acústica, segmentación, constricciones auditivas, constricciones lingüísticas.

The purpose of this work is twofold. On the one hand, we try to establish the main problems of speech perception, on the other, we discuss two possible sources of informational constraints that could lead to the solutions of those problems. The main problem of speech perception derives from the complex relationship between acoustic signal and phonological message recovered by the listener. This problem can be overcome either by using psychophysical constraints imposed by the auditory system or employing specific linguistic knowledge only available when speech sig-

Agradecimientos: Este trabajo ha sido financiado, parcialmente, por los proyectos PR161/93-4791 (UCM) y PB93-0363 (DGICYT). Debo agradecer al Dr. J.E. García-Albea, al Dr. B.S. Rosner y a tres revisores anónimos los comentarios realizados al manuscrito original de este trabajo.

Dirección del autor: Departamento de Psicología Básica I (Procesos Básicos), Universidad Complutense de Madrid, 28223 Madrid. e-mail psbas10@sis.ucmes

nals are to be processed. A third alternative is that speech perception might rest on auditory transformations that are special to speech. It is an empirical question to elucidate what sort of constraints human beings use in the processes of speech perception.

Keywords: Speech Perception, Acoustic Variability, Segmentation, Auditory Constraints, Linguistic Constraints.

La aparición de la comunicación lingüística entre individuos de una misma especie es uno de los logros más sobresalientes de la evolución. Esta especialización biológica dotó a ciertos organismos de dos sorprendentes capacidades. Una de ellas permitía codificar intenciones comunicativas en cierto tipo de señales acústicas. La otra capacidad intervenía en la decodificación del mensaje, esto es, en la recuperación de las intenciones comunicativas de un emisor a partir de la información contenida en las señales acústicas producidas por ese emisor. El estudio de la percepción del habla es el estudio de algunos aspectos de esta última capacidad. El problema general que plantea este estudio es el de encontrar un conjunto de operaciones que permitan derivar, a partir de esas señales acústicas, una representación mental informativa y funcionalmente válida para el procesamiento subsiguiente.

Las operaciones, que pueden entenderse como reglas, serán atribuidas a un sistema de procesamiento que, en principio, permanecerá anónimo. El estímulo de entrada al sistema, la señal acústica, puede ser formalizado como una función de \mathbb{R} en \mathbb{R} , esto es, una función que a un número real —que representa un instante en el tiempo— le asigna otro número real —que representa un nivel de presión sonora—. La respuesta del sistema, su salida, requiere un comentario menos lacónico que los anteriores.

En primer lugar hemos aludido al supuesto de que la salida del sistema ha de ser una representación mental. Esto conlleva un cierto compromiso previo que conviene hacer explícito. Afirmar que un sistema trabaja con representaciones y afirmar que un sistema es intencional, son dos modos de expresar una misma idea. Asumimos la existencia de un sistema intencional en la medida en la que, sólo a través de representaciones, es posible establecer generalizaciones cognitivamente relevantes. Dicho más explícitamente, aunque ciertas propiedades físicas de los estímulos puedan ser considerablemente opacas con relación al papel funcional que cumplen en un sistema de procesamiento, ciertas generalizaciones (no formuladas, necesariamente, en un lenguaje fiscalista) podrían resultar determinantes. Por ejemplo, y referido al caso de la percepción del habla, son muchas las posibles variaciones del estímulo físico que no tienen ninguna consecuencia perceptiva. Que algo sea o no el fonema /p/ depende, claro está, de la constitución física del estímulo; ahora bien, estímulos físicamente bien diferenciables, son desde un punto de vista funcional, o si se prefiere, desde el punto de vista del papel contrastivo que juegan en el sistema, equivalentes. Por tanto, en una primera aproximación, podría considerarse al fonema /p/ como un símbolo capaz de dar cuenta de una generalización que es crítica para explicar el comportamiento de un sistema.

El siguiente paso, obligado, es dotar a las representaciones de un papel causal en la explicación de la conducta. Así llegamos al problema que, de algún modo, ya estaba presente en las ideas de Brentano y que podría ser formulado en forma de interrogante del siguiente modo: ¿cómo es posible que las representaciones mentales entren en el juego causal propio de las entidades físicas? Brentano sólo atisbó la salida dualista a esta coyuntura estableciendo la necesidad de dos causalidades ontológicamente distintas: la que regula los acontecimientos físicos y la que regula la esfera de lo mental. La solución que ofrece la teoría de la mente contemporánea se instala, sin embargo, en una postura netamente materialista. Dicha solución se sustenta en un análisis detallado de los aspectos diversos que constituyen una representación mental. Nunca ha habido gran disputa para reconocer que toda representación mental, como todo símbolo, ha de tener un contenido. El problema estriba en que el contenido semántico de un símbolo no puede ser la causa eficiente del funcionamiento de un sistema. Es claro, sin embargo, que una representación mental no es sólo su contenido semántico. Un análisis riguroso de la noción de representación mental sugiere la necesidad de dos ingredientes adicionales: una forma y una función de instanciación física. La forma de las representaciones propicia la emergencia de una sintaxis que asegura la posibilidad de interacciones causales entre representaciones mentales (transformaciones formales preservadoras del contenido). La instanciación de las mismas en términos de códigos físicos abre las puertas a la interacción causal con el mundo. Se trata, básicamente, de hacer de los códigos simbólicos clases de equivalencia de eventos físicos.

La imagen que surge para caracterizar al sistema de procesamiento es la propia de los sistemas computacionales. La cognición será pues una forma de computación y, por ello, las operaciones a las que aludíamos al formular el problema de la percepción del habla deben ser entendidas como cómputos o cálculos, esto es, como reglas aplicadas a representaciones (*cf.* Pylyshyn, 1984; García-Albea, 1991). Conviene advertir, no obstante, que este planteamiento radical no es unánimemente aceptado por el conjunto de los investigadores de la percepción del habla. Autores surgidos al amparo de la escuela conexionista (*cf.*, por ejemplo, McClelland, Rumelhart y Hinton, 1986; Rumelhart y MacClelland, 1986) han cuestionado la forma clásica de computación y han propuesto un modelo de computación alternativo. En concreto, los sistemas conexionistas se sustentan en simples redes constituidas por una gran cantidad de unidades conectadas entre sí. No obstante, tanto los partidarios de la computación clásica como los defensores del conexionismo, reconocen la existencia de representaciones (Rumelhart y MacClelland, 1986) y, por lo tanto, la afirmación hecha de que percibir habla supone elaborar una determinada representación a partir del estímulo podría ser suscrita por ambas escuelas de pensamiento. Las disensiones entre los partidarios de la computación clásica y los partidarios de la alternativa conexionista radican en el tipo de representaciones y procesos que se atribuyen a la maquinaria mental. En particular, los modelos simbólicos clásicos suponen, como ya ha sido insinuado, que las representaciones mentales poseen una estructura sintáctica y semántica complejas. Los procesos que dichos modelos ejecutan son, por tanto, sensibles a dicha estructura (Fodor y Pylyshyn, 1988). Nada de

esto ocurre en una arquitectura conexionista. Dado que el objetivo del presente artículo no es polemizar acerca de ambos tipos de modelos, aplazaremos hasta un epígrafe posterior la discusión de las implicaciones que tiene la adopción de un modelo conexionista con relación a los problemas de los que se ocupa el presente trabajo.

Hasta el momento hemos definido de modo riguroso el estímulo de la percepción del habla (como una función de \mathbb{R} en \mathbb{R}) y se han hecho algunas consideraciones someras sobre el sistema que lo procesa, lo que nos comprometió con un cierto estilo de buscar explicaciones a los fenómenos. Sin embargo, poco hemos dicho sobre la naturaleza de la representación de salida. A pesar de ello, sí se ha apuntado hacia una restricción que debería cumplir cualquier propuesta que pretenda ser plausible: debe proporcionar una descripción útil al sujeto perceptor. Por tanto y como apunta Marr (1982), la naturaleza específica de la representación tiene que determinarse, por un lado, haciendo referencia a la tarea computacional ejecutada por el sistema (o si se prefiere, refiriéndose al problema de cálculo que el organismo resuelve) y por otro lado, haciendo referencia a las constricciones que operan sobre los procesos posibles (es decir, se trata también de especificar cómo se computa dicha representación). La primera sección de este trabajo se enfrenta al primer aspecto de esta cuestión; la siguiente se ocupará de la segunda vertiente de este asunto.

Los problemas de la percepción del habla

Un acercamiento ingenuo al problema llevaría a la disolución del mismo. Ningún hablante-oyente de un idioma considera que el uso del lenguaje le suponga esfuerzo alguno. Al igual que ocurre en otros campos de la cognición, como la percepción visual, el sujeto resuelve de manera espontánea problemas que, como veremos a renglón seguido, entrañan una gran complejidad. Probablemente, los problemas fundamentales que el organismo resuelve, son los que se derivan de la relación compleja que se da entre la señal acústica y los segmentos fonológicos que se recuperan a partir de ella (Miller y Jusczyk, 1989).

Esta compleja relación se manifiesta, en primer término, en la falta de linealidad característica de la señal de habla: a cada fonema no le corresponde un segmento determinado de la señal (Chomsky y Miller, 1963). De un modo más preciso y más exigente: si al fonema X le sigue el fonema Y en la representación fonológica, debe haber una porción de la señal acústica asociada al fonema X que debe preceder a la porción de señal acústica asociada al fonema Y (Pisoni y Luce, 1987). Como veremos unos párrafos más adelante cuando tratemos el problema de la segmentación, este requisito de linealidad es claramente violado.

Pero quizá el problema que ha resultado más esquivo y, por tanto, el que ha de imponer más severas constricciones sobre las operaciones que han de ser ejecutadas sobre la señal de entrada, es el de la variabilidad. Este problema estriba en que no existe un conjunto invariante de características acústicas que proporcionen información necesaria y suficiente para señalar la presencia de un seg-

mento fonológico (o incluso de un rasgo distintivo) (Miller y Jusczyk, 1989; Chomsky y Miller, 1963; Lindblom, 1986). Ningún conjunto de características acústicas es necesario puesto que conjuntos diferentes pueden señalar la presencia de un mismo segmento fonológico. Ningún conjunto de características acústicas es suficiente porque, en distintos contextos, un mismo conjunto de características puede señalar la presencia de segmentos fonológicos distintos (Lieberman, 1982; Repp, 1982). Veamos algún ejemplo que ilustre estos comentarios.

Lieberman, Delattre, Cooper y Gerstman (1954) estudiaron las claves acústicas que señalan el lugar de articulación en las consonantes sonoras (/b/, /d/ y /g/). Sin entrar en detalles, se observó una clara relación entre lugar de articulación y las transiciones del segundo formante (F2). La distinción acústica podía referirse tanto a la magnitud como a la dirección de la transición. La sílaba /bi/ se caracteriza por una transición del segundo formante ascendente. Lo mismo ocurre en la sílaba /di/ salvo que dicha transición es más breve. Por el contrario para la sílaba /gi/ F2 presenta una clara transición descendente. Si existieran propiedades acústicas invariantes que marcaran los distintos contrastes fonológicos, cabría esperar que las claves encontradas seguirían siendo funcionalmente válidas para el resto de los contextos vocálicos. Es decir, si deseo sintetizar, por ejemplo, la sílaba /da/, debería ser suficiente con definir los formantes apropiados que caracterizan a la vocal /a/ y añadir la clave acústica descubierta para la /d/, esto es, una transición ascendente en F2. Desgraciadamente, un estímulo físico tal produciría un percepto muy cercano a la sílaba /ba/ (cfr. Delattre, Lieberman y Cooper, 1955). Parece claro que no existe información acústica invariante (no dependiente del contexto acústico vecino, por ejemplo) que permita asegurar que estamos ante uno u otro segmento fonológico. Por resumirlo siguiendo la misma lógica que utilizamos al principio: no es necesaria la presencia de una transición ascendente breve para que se perciba una /d/ (puesto que con transiciones descendentes se perciben sílabas como /da/, /do/ y /du/) ni tampoco es suficiente, ya que esas mismas transiciones ascendentes breves provocan la percepción de las sílabas /ba/, /bo/ y /bu/.

No es de extrañar que estos dos problemas, el de la linealidad y la variabilidad, desemboquen, más por acumulación que por implicación, en un tercero: el de la segmentación. El problema consiste en la imposibilidad de fraccionar la señal de habla en unidades autónomas que correspondan a secuencias de fonemas (o de sílabas, o de palabras) (Lindblom, 1986). Aunque, para ser precisos, habría que apuntar que la segmentación de la señal de habla sí es posible atendiendo a criterios estrictamente acústicos (Fant, 1962, 1973). El problema es que el número de segmentos acústicos es típicamente mayor que el número de segmentos fonológicos. La formulación más expresiva de este problema se debe a Hockett (1958) quien señaló que, más que a las cuentas de un collar, la instanciación física de los fonemas se parecía a «una fila de huevos aplastados por un rodillo». El problema de la segmentación es una consecuencia de la coarticulación que se da en la producción del habla. Cabe preguntarse cómo es posible que un fenómeno que provoca tamaños problemas puede tener éxito adaptativo en el decurso de la evolución. La respuesta es que dicho mecanismo —el de la coarticulación— proporciona una velocidad de transmisión de segmentos fonémicos

mucho mayor de la que se derivaría del procedimiento de asignar un elemento fonémico a cada gesto articulatorio. Aunque la tasa de habla puede ser muy variable, se pueden tener velocidades de transmisión del orden de 20-25 fonos/s. Nada indica que se puedan articular 20-25 segmentos acústicos/s, pero aun si esto fuera posible, el oyente sólo sería capaz de percibir zumbidos o tonos amalgamados (Liberman y Studdert-Kennedy, 1978).

El problema de la segmentación se manifiesta en sus dos formas lógicamente posibles. Por una parte, un mismo segmento acústico contiene, habitualmente, información acerca de más de un segmento fonológico (es decir, se produce una transmisión en paralelo de información concerniente a la identidad fonológica). Por otra parte, la información referida a un mismo segmento fonológico se encuentra distribuida en distintos segmentos acústicos (es decir, deben existir procesos de integración de información distribuida en el tiempo) (Fant, 1962; Liberman, Cooper, Shankweiler y Studdert-Kennedy, 1967). De nuevo, es fácil encontrar ejemplos que ilustren este problema.

Yeni-Komshian y Soli (1981) emplearon sílabas compuestas por una consonante fricativa y una vocal en un experimento perceptivo. En particular, utilizaron 4 tipos de fricativas emparejadas con tres vocales distintas. En este tipo de estímulos es relativamente simple determinar la parte que corresponde a la fricación (porción no armónica) y la parte periódica que corresponde a la vocal. El resultado general que obtienen es que los segmentos fricativos y vocálicos determinados sobre la base de criterios acústicos, no se corresponden de manera directa con los segmentos fonológicos percibidos. Básicamente, el experimento consistió en desgajar la porción vocálica de la fricativa y presentar esta última a los sujetos para su identificación. Los sujetos no sólo eran capaces de determinar qué fricativa se les había presentado, sino también, hasta cierto punto, qué vocal había estado asociada a esa porción. Este experimento pone de manifiesto las dos caras del problema de la segmentación: muestra que un mismo segmento acústico (el de fricación) transmite información en paralelo acerca de dos segmentos fonológicos distintos y además (dado que la información primigenia para identificar la vocal está en la porción periódica) muestra que dos segmentos acústicos contienen información acerca de un mismo segmento fonológico (Miller y Juszcyk, 1989).

No son éstos los únicos problemas de los que el organismo ha de hacerse cargo si pretende ser un procesador eficaz de señales de habla. Consideremos, por ejemplo, los problemas de normalización temporal. Es sabido que la duración de un segmento puede constituirse en una clave indicadora de la presencia de un determinado fonema (como es el caso de la cualidad de las vocales inglesas) (véase Miller, 1981 para una revisión del tema). Ahora bien ¿cómo distinguir estos efectos de la duración de otros que son irrelevantes y que son dependientes, por ejemplo, de la tasa de habla? También se podría aludir a los problemas de normalización del hablante. Distintos hablantes difieren en la longitud y forma de sus trectos vocales, en sus estrategias articulatorias y coarticulatorias y en el dialecto que utilizan (Stevens, 1972a). No es necesario extenderse en éstos y otros problemas asociados (como los derivados de ambientes más o menos ruidosos, con más o menos reverberación o los asociados al canal

de comunicación utilizado) porque en definitiva desembocan en los grandes problemas que hemos desarrollado con algo más de detenimiento.

Merece la pena recordar, no obstante, la posibilidad de que los procesos que se encargan de la segmentación del *input* no sean independientes y anteriores a los de etiquetado fonológico (Klatt, 1979). En este caso la segmentación del *input* sería, en cierto sentido, un epifenómeno: no es más que el resultado automático de un proceso de decisión fonológica. Digo sólo en cierto sentido, porque seguiría existiendo el problema de delimitar los segmentos de la señal sobre los que se han de tomar las decisiones. En cualquier caso, parece ser que los criterios de segmentación han de estar apoyados en conocimiento que no proporciona únicamente el estímulo, por ejemplo, conocimiento de las relaciones articulatorio-acústicas o conocimiento de las restricciones fonológicas del idioma (Klatt, 1978; Cole, Rudnicki, Zue y Reddy, 1980). Como se puede apreciar estas consideraciones nos introducen en un nuevo debate con una doble vertiente. Por un lado, se plantea la polémica sobre la existencia o no de unidades intermedias entre la señal acústica y el léxico; por otro lado, si es que estas unidades existen, se abre el interrogante acerca de su naturaleza o, en otras palabras, se plantea el problema de la unidad de procesamiento. Volveremos sobre este tema en un epígrafe posterior para no romper aquí el argumento que está siendo presentado. Retomemos, pues, el hilo conductor de este apartado.

En el intento de determinar la naturaleza de la representación de salida, hemos discutido, en primer lugar y de modo informal, algunos de los problemas a los que el organismo se enfrenta para poder recuperar un mensaje a partir de una señal acústica. Resumiendo, se podría decir que estos problemas son dos: falta de invarianza y de segmentación del estímulo; el estímulo no se nos presenta ni invariante, ni apropiadamente segmentado con respecto al mensaje lingüístico que se pretende recuperar. Como ya ha quedado dicho, la representación de salida tiene que proporcionar información útil al organismo. Ahora podemos afirmar que, para cumplir este objetivo, resulta imprescindible contar con una representación invariante y segmentada. Invariante con respecto a factores de variabilidad que son irrelevantes y segmentada en la medida en que ha de constituirse en un símbolo atómico. Pero, ¿qué tipo de sistema y qué tipo de recursos pueden ser capaces de ajustarse a tales requerimientos? Vamos a tratar, en unos retazos impresionistas, de dar el siguiente paso que nos acerca a las respuestas de estas preguntas.

Una de las propiedades fundamentales del lenguaje es su naturaleza cuántica (Lindblom, 1986; Stevens, 1972b; 1989). Es decir, el lenguaje es, esencialmente, un sistema reglado constituido por unidades discretas. El mismo Halle (1964) reconocía este hecho al hacer de la naturaleza segmental del habla la clave de los desarrollos de la lingüística contemporánea (aunque no todos están de acuerdo con este punto; véase, por ejemplo, Moll, Zimmerman y Smith, 1976; Lindau, 1978). La existencia de reglas y unidades es lo que dota al lenguaje de una de sus características más destacadas: la «compositividad», característica de la que dependen otras dos propiedades esenciales del lenguaje como son su productividad y su sistematicidad (Fodor y Pylyshyn, 1988; García-Albea, 1992). Las unidades pueden ser combinadas; las reglas indican cómo (Ja-

kobson, Fant y Halle, 1951; Chomsky y Halle, 1968). Un sistema capaz de representar propiedades abstractas relativas a las secuencias de sonidos que constituyen el habla, es un sistema que, al menos en principio, es capaz de hacerse cargo de los problemas de invarianza y segmentación. Lo que se precisa, por tanto, es un sistema sensible a propiedades estructurales que no se identifican con —aunque se apoyen en— eventos de carácter físico. En otras palabras, se requiere un sistema intencional.

Los comentarios que acaban de ser expuestos con respecto a las propiedades básicas del lenguaje (su carácter cuántico, compositivo y sensible a aspectos formales) inspiran una reflexión inmediata, una reflexión crucial para responder la pregunta acerca de los recursos con los que dotar al sistema de procesamiento. El lenguaje no es sólo el objeto externo que el sujeto ha de percibir, sino el tipo de conocimiento (tácito) que el sujeto pone en juego para percibirlo. Reflexión poco original, dicho sea de paso (véase, por ejemplo, Chomsky, 1980), pero a la vez piedra angular de la investigación psicolingüística de los últimos 30 años. Aunque no es el momento de hacer una defensa de la «realidad psicológica» de las representaciones fonológicas, es preciso, al menos, señalar las distintas fuentes de evidencia que se han manejado en este sentido. Existen argumentos de carácter estrictamente formal, es decir, derivados del análisis de las formas lingüísticas (datos distribucionales sincrónicos dentro de una lengua, datos acerca de la verosimilitud de determinados cambios en las formas fonológicas a través del tiempo, datos acerca de la mayor verosimilitud de que se produzcan determinadas situaciones fonológicas a través de los idiomas y, en general, el mismo hecho de que se den fenómenos de carácter estrictamente fonológico (Nearey, 1981; Kenstowicz y Kisseberth, 1979). También existen argumentos de carácter empírico. Por ejemplo en el campo de los errores espontáneos e inducidos en la producción del habla (Fromkin, 1980), en el de los errores producidos por interferencia en la memoria a corto plazo (Wang y Bilger, 1973), incluso observando la actuación de hablantes que desconocen sistemas ortográficos (Sapir, 1963; Read, 1971).

Las restricciones computacionales en los procesos de la percepción del habla

Hasta ahora se ha tratado con algún detalle el tema de los problemas de cálculo, vamos a introducir el referido a las constricciones que operan sobre los procesos posibles. Este paso es obligado por varias razones. Se podría empezar el argumento, a la antigua usanza, estudiando las relaciones entre competencia y actuación. Precisamente la competencia viene a ser ese conocimiento tácito, al que aludíamos más arriba, y del que el sujeto hace uso en su actividad lingüística. Lo que tratamos de justificar a continuación es que no basta con la descripción estructural de ese conocimiento para dar cuenta de la actuación.

La búsqueda de constricciones que definan procesos es obligada debido, entre otras cosas, a lo que se ha denominado «condición de prevalencia de la es-

estructura» (Valian, 1979). Dicha condición asume que aunque las representaciones que propone la teoría de la competencia lingüística son agentes estructurales de la actividad lingüística, los procesos que deben ser postulados para dar cuenta de la actuación no tienen por qué ser isomórficos con respecto a las reglas de la competencia. Estas consideraciones, aunque habitualmente ejemplificadas en el ámbito del procesamiento sintáctico, pueden ser trasladadas a otros ámbitos, y fuerzan precisamente esa búsqueda de constricciones que singularizan al proceso adecuado. Marr (1982) enfatiza la importancia capital de encontrar constricciones de carácter físico (propiedades que se imponen de modo natural a partir de la estructura del mundo) para determinar soluciones únicas a los problemas de procesamiento. Una constricción, en términos más generales, podría ser considerada como un tipo de información adicional capaz de restringir el número de posibles soluciones en el espacio específico de un problema. En este sentido, se podría hablar, no sólo de constricciones físicas, sino de otros tipos de constricciones que vendrían a restringir tanto el número como la naturaleza de los procesos posibles.

Dos tipos generales de constricciones se han venido discutiendo en la literatura como fundamentales a la hora de entender la naturaleza de los procesos de la percepción del habla. De un lado se ha recurrido a las reglas de funcionamiento del sistema auditivo, de otro, a las reglas propias del sistema de conocimiento lingüístico que se supone utilizan los hablantes-oyentes de un idioma.

La importancia de las constricciones impuestas por el sistema auditivo en el procesamiento del habla han sido resaltadas por diversos autores. Se ha llegado a sugerir que este tipo de constricciones podrían determinar el tipo de distinciones acústicas que usan los distintos lenguajes para formar contrastes fonológicos (Stevens, 1981). Esto ha motivado una extensa investigación acerca de las propiedades básicas de las estructuras y los procesos del sistema auditivo. Este tipo de investigación se ha abordado desde dos ángulos distintos. Por un lado, ha habido una extensa investigación fisiológica tratando de determinar los principios de funcionamiento del sistema auditivo (Delgutte y Kiang, 1984; Sachs y Young, 1979; Young y Sachs, 1981, Sinex y Geisler, 1983). Por otro lado, el mismo problema ha sido atacado desde su vertiente psicofísica (Zwicker, Terhardt y Paulus, 1979; Searle, Jacobson y Rayment, 1979; Klatt, 1982; Kewley-Port, 1983). Sin entrar en detalles, podemos señalar que estos estudios han proporcionado constricciones acerca de la resolución temporal y espectral del sistema (forma de los canales y tamaño de la ventana de análisis), acerca de las transformaciones que hay que efectuar sobre los valores de la intensidad del estímulo, introducción de no linealidades en el sistema, introducción de mecanismos de supresión lateral para enfatizar determinados picos espectrales etc. Aunque, como se puede apreciar, la mayoría de estos trabajos han centrado su atención en los sistemas más periféricos (cóclea y nervio auditivo), se empieza a reconocer la necesidad de tomar en cuenta los componentes más centrales del sistema (Watson y Foyle, 1985).

La idea de fondo de este acercamiento psicoacústico a los problemas de la percepción del habla no es sólo que los segmentos fonémicos que pertenecen a una misma clase tienen propiedades acústicas comunes. La versión rigurosa de

esta hipótesis afirma que son estas propiedades acústicas las que se utilizan en el procesamiento *on-line* de la señal de habla. Es decir, el procesamiento de la señal de habla supondría la identificación de estas propiedades acústicas de la señal tal y como son representadas por el sistema auditivo; la cuestión crítica es, por tanto, cuáles son las distinciones acústicas relevantes y cómo son tratadas por el sistema auditivo.

Para ilustrar este punto consideremos el papel que juega el sistema auditivo en el procesamiento de las claves acústicas que proporcionan las transiciones frecuenciales propias de los estímulos lingüísticos. Está fuera de toda duda el hecho de que tanto la duración como la dirección de las transiciones de los formantes proporcionan claves acústicas para contrastes lingüísticos como el modo y el lugar de articulación. Por ejemplo, si una determinada transición (entre las frecuencias apropiadas) dura 60 ms. o más el sujeto percibirá la sílaba /wa/, pero si la duración es inferior a esos 60 ms. el sujeto percibirá la sílaba /ba/. En segundo lugar, la dirección de transiciones rápidas (entre 40 y 60 ms.) sirve para diferenciar, por ejemplo, las sílabas /ba/, /da/ y /ga/ (cfr. Potter, Knopp y Knopp, 1966). Parece que las duraciones de 40 a 60 ms. tienen una especial significación para efectuar contrastes lingüísticos. La pregunta que queda abierta es si esto es así debido al modo en que el sistema auditivo trata estas duraciones en el contexto de las transiciones frecuenciales. Dicho de otro modo, ¿provoca el procesador auditivo una mayor sensibilidad natural para las transiciones de formantes que transcurren entre los 40 y los 60 ms.? Jamieson y Slawinska (1983; 1984) realizaron varios experimentos con estímulos no lingüísticos (formados por una transición frecuencial y una parte estacionaria) tratando de averiguar si los seres humanos venían equipados con dispositivos que favorecían la percepción de este tipo de transiciones frecuenciales (para un resumen de estos trabajos véase Jamieson, 1987). En todos los experimentos los sujetos mostraron funciones no monotónicas de discriminación con máximos situados en el rango de 40-60 ms. confirmando la idea de que para esa región el sistema auditivo posee una mayor sensibilidad natural. Ésta sería la razón por la que distintas lenguas tenderían a contrastar categorías fonológicas apoyándose en la información acústica contenida en cambios frecuenciales con duraciones en torno a los 50 ms. Si dichas variaciones en frecuencia duraran más o menos tiempo, las claves acústicas basadas en estas variaciones serían más difícilmente discriminables. Por tanto, la distinción /ba/-/wa/ o las distinciones /ba/-/da/-/ga/ se derivan de factores psicoacústicos que provocan ciertas discontinuidades sensoriales.

En otra clase natural pueden reunirse otro grupo de constricciones que se refieren de modo específico a aspectos del sistema lingüístico y no del sistema auditivo en general. Se podría hablar, en primer lugar, de las constricciones articulatorias impuestas a la representación del habla (Ohala, 1981). Así, factores tales como la elasticidad del tejido de los articuladores, las conexiones anatómicas entre articuladores, propiedades aerodinámicas del sistema etc., determinan algunas de las características que estarán presentes en la señal de habla. Por ejemplo, se ha demostrado que el valor del VOT (tiempo de inicio de la sonoridad) en oclusivas sordas varía en función del contexto vocálico. En particular, ante vo-

cales altas el VOT es mayor que ante vocales bajas (Summerfield, 1975). Dado que las vocales altas ofrecen una mayor resistencia al aire que sale de la cavidad oral, se ha sugerido (Ohala, 1976) que este hecho retrasa el momento en que se produce la suficiente presión transglotal para iniciar la sonoridad, y por tanto, el VOT se alarga. Efectos semejantes referidos a alteraciones en la frecuencia fundamental de segmentos vocálicos pueden encontrarse en Hombert, Ohala y Ewan, 1979. Por volver a un ejemplo conocido, el conocimiento de constricciones articulatorias explica por qué las distintas realizaciones acústicas de /d/ en /di/ y /du/ no son obstáculo para obtener un percepto invariante. Por ser la /d/ una consonante dental (alveolar en inglés), al inicio de /di/ o /du/ la lengua se sitúa en la zona de los dientes. Por tanto, la configuración del tracto vocal y, por ende, las frecuencias de resonancia, son similares. El resultado es que la frecuencia de inicio del segundo formante (F2) es similar. ¿Por qué se producen las diferencias en la transición de los formantes? Sencillamente porque la frecuencia de resonancia que corresponde a F2 es mayor en el caso de la /i/ que en el caso de la /u/. Por ello, desde la posición inicial, al articular /di/ nos encontramos con una transición ascendente, y al articular /du/ con una descendente. El aspecto importante de la cuestión es que, aun resultando dos estímulos físicamente distintos, el gesto articulatorio ha sido el mismo (oclusión del tracto vocal en un punto determinado) y las diferencias físicas son producto de constricciones articulatorias susceptibles de ser regladas.

Otro tipo de constricciones a considerar dentro de este mismo grupo, son las que se refieren a los aspectos suprasegmentales de la señales de habla. Factores tales como la entonación o la velocidad de habla son causantes, como ya comentamos, de algunos de los problemas de variabilidad que afectan al estímulo. Por tanto, como muy bien señala Nootboom (1981), estudiar los efectos de factores suprasegmentales en la percepción de segmentos fonológicos invariantes, puede ayudar a determinar cómo de hecho se extraen esos segmentos invariantes a partir de señales variables. Es sabido, por ejemplo, que el ritmo y la entonación de las frases afectan a la estructura segmental debido a los efectos que ocasionan en la estructura silábica (Allen, 1975; 1981). (Para una revisión de las restricciones fonotácticas que afectan a la distribución de los segmentos en el marco de las sílabas, véase Bell y Hooper, 1978).

Un tercer tipo de constricciones genuinamente lingüísticas son las acuñadas en las reglas de la fonología segmental. En primer lugar, convendría señalar que no cualquier sonido posible es un sonido que pertenezca, de hecho, a algún lenguaje natural (Greenberg, 1963, 1966). Parece ser que existe un inventario finito de sonidos de habla y que, con unas 12 características distintivas, es posible definir cualquiera de los sonidos que forman parte de cualquier lenguaje natural conocido (Jakobson y Halle, 1956). A partir de aquí es posible definir reglas fonológicas que podrían entenderse como constricciones que se imponen a las secuencias de segmentos y que determinan la «gramaticalidad» de los mismos. Por ejemplo, la forma que adopta el morfema de plural en inglés depende de las características del último segmento del nombre en cuestión (Halle y Stevens, 1979). En particular, el sufijo de plural sería [iz] si el nombre termina con una consonante coronal estridente, en otro caso es [s] si el nombre acaba en una

consonante sorda y en cualquier otro caso es [z]. Ejemplos parecidos pueden encontrarse en castellano. Por ejemplo, la realización fonética de las oclusivas sonoras depende de la posición en la que se encuentre el segmento en cuestión. Tanto el fonema /b/ como el fonema /d/ y el fonema /g/ tienen dos alófonos, uno oclusivo y otro fricativo. El alófono oclusivo aparece cuando el segmento ocurre tras una pausa o tras una consonante nasal (en el caso de la /d/ también tras una lateral), en el resto de contornos se instancia el alófono fricativo (Quilis y Fernández, 1992).

En contraposición con las propuestas de la alternativa psicoacústica, la concepción que acabamos de exponer con relación al procesamiento del habla, no considera primordial el estudio de las representaciones que proporciona el sistema auditivo como tal. Lo que resulta crucial es el conjunto de constricciones lingüísticas que se imponen en el curso del procesamiento de la señal. Podríamos decir que lo que el sujeto «oye» cuando atiende a una señal de habla es lo que determinan las reglas fonológicas que tiene internalizadas (Chomsky y Halle, 1968). En este sentido ha argumentado recientemente Kaye (1989). La pregunta de partida sería la siguiente: ¿por qué encontramos generalizaciones fonológicas susceptibles de ser regladas en todos los sistemas lingüísticos estudiados? Una primera respuesta a esta pregunta, que nos alejaría de la interpretación «psicologista» que está siendo expuesta, acudiría a la conocida ley del mínimo esfuerzo. Esto es, la concatenación de morfemas para formar palabras puede producir secuencias «incómodas», secuencias que requieran un mayor gasto energético al sistema articulador. Por esta razón, los procesos fonológicos tenderían a eliminar estas secuencias incómodas (variando las características fonéticas de los segmentos en cuestión). Así podríamos explicar, por ejemplo, la regla con la que concluía el párrafo anterior: el plural de la palabra inglesa «dog» es dog[z] y no dog[s] sencillamente porque esta última expresión es difícilmente pronunciable. Sin embargo, estas explicaciones no parecen ser suficientemente satisfactorias. Por ejemplo, es un hecho conocido que ciertas secuencias de sonidos modificados por un proceso fonológico o de cambio histórico resultan ser el resultado final de otros cambios o procesos (cfr. Kaye, 1989). Por ejemplo, la palabra latina «doctor» pasa a ser «dottore» en italiano. Presumiblemente para deshacerse de la incómoda secuencia de las dos oclusivas -kt- y reducirla a un sólo gesto articulatorio -t-. El problema es que en el paso del árabe clásico al árabe actual que se habla en Marruecos encontramos el fenómeno contrario. La forma «kataba» pasa a ser «ktib». Parece, pues, que la existencia de procesos fonológicos es útil en la medida en que posibilitan el procesamiento del habla de la manera eficaz que conocemos. Como hemos comentado, resolver el problema de la segmentación es un paso crucial para poder construir procesadores de habla eficaces. Este problema se extiende más allá de la segmentación de segmentos fonológicos y afecta también a la segmentación de las piezas léxicas. En el habla fluida es difícil determinar dónde acaba una palabra y empieza la siguiente. Existen procesos fonológicos que sirven al propósito de marcar estas fronteras. Los sistemas de acentuación son claros candidatos. Por ejemplo, en francés todas las sílabas que son final de palabra llevan acento. En húngaro son las sílabas iniciales las que lo llevan. En ambos casos el conocer esta informa-

ción acerca del sistema de acentuación es crítico a la hora de determinar dónde se encuentran las fronteras entre palabras.

Finalmente, resulta interesante reflexionar acerca de una última posibilidad. Podría ocurrir que existieran procesos auditivos que sólo se activaran ante señales de habla. O dicho desde otra perspectiva, podría ocurrir que la investigación más reciente estuviera encontrando los mecanismos auditivos que implementan (que hacen causalmente eficientes) las constricciones a que se ve sometido el procesamiento del habla. En este sentido, es muy ilustrativo el modelo reciente de percepción de vocales propuesto por Rosner y Pickering (1994), así como ciertos resultados experimentales encontrados en este mismo ámbito de estudio (López-Bascuas, Fahey, Hernández y Gómez, 1995; Fahey y López-Bascuas, 1994). Aunque estos autores califican su teoría como «auditiva», uno de sus componentes es explícitamente definido como específico para el procesamiento del habla.

En concreto, Rosner y Pickering (1994) proponen cinco fases en el análisis de la señal lingüística. En primer lugar, se realiza una transformación sobre el eje frecuencial del espectro con el fin de obtener una representación psicológicamente relevante de las frecuencias que componen la vocal. En segundo lugar, este espectro transformado es procesado por un conjunto de filtros que respetan ciertas constricciones psicoacústicas conocidas (Glasberg y Moore, 1990). Esta operación genera un patrón auditivo de excitación. En tercer lugar, se incluyen en el modelo los efectos de supresión. El cuarto componente del modelo transforma los valores de intensidad del espectro modificado en términos de la sonoridad percibida. El resultado final es un patrón auditivo de densidad sonora o, como sus autores lo denominan, un ALDP. Todos los sonidos, sean de habla o no, generan ALDPs. Ahora bien, el quinto componente del modelo es estrictamente fonético y, por tanto, sólo afecta a las señales lingüísticas. La introducción de este componente está motivada por las características de procesamiento del sistema auditivo, características que son recogidas por los filtros utilizados en el modelo. La anchura de banda de estos filtros aumenta según incrementamos la frecuencia central del filtro con lo que los armónicos superiores de la vocal resultan cada vez más difíciles de resolver. Esto hace que la identificación de formantes sea difícil en la zona de baja frecuencia del espectro. De este modo, la identificación del primer formante sería difícil para las vocales anteriores; para las vocales posteriores incluso podría ser problemática la localización del segundo formante. La solución propuesta por Rosner y Pickering (1994) supone el concurso de una segunda ventana de integración. Esta segunda ventana suavizaría los máximos presentes en el ALDP debidos a armónicos próximos, proporcionando un único máximo local en la región pertinente del espectro. Por tanto, la aplicación de esta segunda ventana de integración al ALDP genera como resultado un patrón fonético de densidad sonora (PLDP). Los detalles acerca del funcionamiento de esta segunda ventana de integración no son relevantes para este trabajo, lo esencial es el hecho de que esa segunda integración es un operación auditiva que sólo sería relevante para el tratamiento de señales de habla.

Por su parte, López-Bascuas *et al.* (1995) y Fahey y López-Bascuas (1994) también discuten esta posibilidad en el contexto más restringido de la

percepción y la producción de la altura vocálica. En particular, ambos trabajos ponen a prueba la hipótesis de que la distinción [+/- alta] pudiera estar basada en una discontinuidad psicoacústica que aparece en torno a un límite de integración de 3-3.5 Bark. La escala de Bark permite transformar valores frecuenciales en valores tonales psicológicamente relevantes. De forma un tanto imprecisa, pero comprensible, podríamos decir que 1 Bark equivale a la anchura de una banda crítica. Hoemecke y Diehl (1994) encuentran evidencia positiva utilizando vocales inglesas y una muestra de sujetos americanos hablantes de inglés. Sin embargo, Fahey y López-Bascuas (1994) y López-Bascuas *et al.* (1995) obtienen resultados negativos empleando vocales españolas y una muestra de sujetos españoles. Estos autores sugieren la posibilidad de que el integrador de frecuencias en torno a 3-3.5 Bark podría ser (en el caso de que existiera) específico para el habla y dependiente del idioma.

Conclusiones

Este trabajo ha intentado presentar una polémica actual partiendo de unos supuestos teóricos claramente definidos. Sin embargo, no es posible obviar otros marcos teóricos que, de un modo u otro, afectan a los problemas que aquí han sido tratados. En primer lugar, es preciso reconocer que los modelos conexionistas han arrojado una luz nueva e interesante sobre los problemas clásicos de la percepción del habla. El modelo TRACE (McClelland y Elman, 1986) proporciona un claro ejemplo a este respecto. TRACE es un modelo de activación interactiva jerarquizado en tres niveles de representación: nivel de rasgos, nivel de fonemas y nivel de palabras. Los problemas de variabilidad del *input* afectarían a una arquitectura clásica tipo von Neumann, sin embargo, en una arquitectura PDP, la variabilidad del *input*, en tanto en cuanto representa información sobre interacciones relevantes, puede ser utilizada para facilitar las tareas perceptivas del sistema (Elman, 1989). Por otro lado, TRACE no precisa realizar una segmentación manifiesta de la señal y, sin embargo, obtiene representaciones segmentales del *input*. Como el mismo Elman (1989) señala, la segmentación resulta ser un resultado de la actividad perceptiva y no una condición previa para que ésta pueda llevarse a efecto. Aunque TRACE utiliza representaciones locales, es posible diseñar sistemas que generen por sí mismos un conjunto de representaciones distribuidas. Para ello basta con organizar una red con una capa oculta y utilizar un algoritmo de retropropagación (*back-propagation*) para entrenar la red. En estas condiciones, es habitual que la red consiga reconocer palabras y generar ciertas representaciones subléxicas que aparecen distribuidas en el patrón de activación de las unidades de la capa oculta. Esto indicaría que tales representaciones son productos de la actividad del sistema y no conocimiento explícitamente representado a priori en dicho sistema. A pesar de todo, TRACE no está exento de problemas. Aparte de su arquitectura altamente implausible (la red léxica se replica en numerosas ocasiones para permitir un reconocimiento de palabras independiente del tiempo), TRACE ha sido criticado

por su carácter interactivo (Norris, 1994). Trabajos recientes parecen mostrar el carácter autónomo de los procesos de identificación fonémica (Cutler, Mehler, Norris y Segui, 1987; Frauenfelder, Segui y Dijkstra, 1990; McQueen, 1991). Sin embargo, esta crítica sólo afecta al carácter interactivo de TRACE, no a su arquitectura conexionista. De hecho, es posible construir redes conexionistas donde el flujo de información es enteramente guiado por los datos (*bottom-up*) y que, sin embargo, dan cuenta de algunos fenómenos habitualmente atribuidos al procesamiento guiado conceptualmente (*top-down*) (Norris, 1992; 1993; 1994). Aunque no es éste el lugar adecuado para presentar la polémica con el conexionismo, conviene recordar que sus presupuestos teóricos han sido severamente criticados (Fodor y Pylyshyn, 1988).

Otros marcos teóricos también han tratado de eludir los problemas de variabilidad y segmentación del estímulo. El modelo LAFS (*Lexical Access From Spectra*) de Klatt (1979; 1989) proporciona un claro ejemplo. En LAFS los emparejamientos léxicos se hacen sobre la base de secuencias espectrales desprovistas de cualquier tipo de etiqueta lingüística. El léxico se supone constituido por una red que representa secuencias de difonos a los que se asocia un patrón espectral determinado. Reconocer una palabra simplemente supone recorrer un determinado camino en la red de decodificación. De este modo, no se asume ningún tipo de segmentación previa del *input* (ni en términos de características ni en términos de fonemas) y tampoco se asume invariabilidad en la señal. El sistema procede computando espectros y buscando el patrón correspondiente en la red de decodificación. Sin embargo, esto no quiere decir que el sistema sea ciego a restricciones lingüísticas. En concreto, los segmentos fonéticos y las reglas fonológicas cumplen una misión importante en el proceso de codificación de la red. Además, no se resuelven de hecho todos los problemas de variabilidad de la señal. Por ejemplo, es difícil determinar un preprocesamiento eficaz para normalizar variaciones debidas a diferencias en la longitud del tracto vocal del hablante. Finalmente, como ocurre con TRACE, la arquitectura general del sistema es psicológicamente implausible (*cfr.* Klatt, 1989).

Otra propuesta semejante a la anterior, pero quizá no tan radical, puede encontrarse en un trabajo reciente de Marslen-Wilson y Warren (1994). Estos autores se preguntan acerca de cuál es la naturaleza del *input* que ha de contactar con el léxico interno. Marslen-Wilson y Warren (1994) se decantan finalmente por una teoría de acceso directo donde no existe una computación explícita de unidades fonémicas; por el contrario, las características distintivas extraídas de la señal se emparejarían directamente con representaciones léxicas. Por su parte, las entradas léxicas estarían organizadas en términos de esas características distintivas (Lahiri y Marslen-Wilson, 1991; 1992). El entorno general de procesamiento lo proporcionaría el marco conceptual recogido en las últimas versiones del modelo de la Cohorte (Marslen-Wilson, 1993). Sin embargo, en un trabajo reciente (Radeau, Morais y Segui, 1995) se invita a pensar en la posibilidad de que existan representaciones preléxicas en el proceso de reconocimiento de palabras. Estas consideraciones nos introducen en otro terreno ampliamente debatido, el que se refiere a la unidad de procesamiento del habla. Las propuestas en este ámbito han sido variadas. Elman (1989) menciona hasta

ocho unidades distintas que podrían ser relevantes: los patrones espectrales, los rasgos, los difonos, los alófonos, los fonemas, las demisílabas, las sílabas y los morfemas. La situación se torna incluso más compleja cuando se considera que las posibles estrategias de segmentación de la señal podrían verse afectadas por la lengua materna de los sujetos y por el tipo de tarea utilizada en los experimentos (Cutler, Mehler, Norris y Seguí, 1986; Sebastián-Gallés, Dupoux, Seguí y Mehler, 1992). En cualquier caso, todas estas últimas propuestas no hacen sino poner de manifiesto lo intrincados que resultan los problemas de variabilidad y segmentación de la señal del habla. De hecho, la búsqueda de representaciones intermedias entre la señal y el léxico interno es un intento de tratar la variación acústica acudiendo a unidades (generalmente lingüísticas o semejantes a ellas) que proporcionen un entorno más estable de procesamiento. En este sentido, toda esta línea de investigación constituye un intento de responder (parcialmente) a los problemas que hemos venido considerando centrales en el estudio del procesamiento del habla.

En definitiva, empezamos definiendo el problema de la percepción del habla y, de algún modo, este trabajo se ha conformado alrededor del análisis de los términos que intervinían en esa definición. Percibir habla supone convertir un estímulo físico en una representación mental. Para ello se requiere un sistema de procesamiento. No hubo dificultad en definir el estímulo de entrada a dicho sistema. Aunque más susceptible de discusión, tratamos de defender la idea de que se precisa un sistema intencional para procesar dicho estímulo, es decir, un sistema capaz de manejar símbolos. De hecho, lo único que indicamos, en un primer momento, con respecto a la representación de salida, es que debía proporcionar una descripción simbólica útil para el sistema. Por tanto, la siguiente tarea consistió en tratar de determinar la naturaleza de dicha representación y tratar de buscar constricciones a los procesos posibles. Esto nos llevó a explicitar los problemas de cálculo con los que se enfrenta el sistema (linealidad, invarianza y segmentación) y a distinguir entre dos tipos generales de constricciones (auditivas y lingüísticas). La primera tarea culminó con la propuesta de una representación de carácter discreto constituida por propiedades invariantes. La segunda tarea abre un interrogante fundamental para la teorización en el ámbito de la percepción del habla: ¿qué tipos de constricciones son los que operan para obtener dicha representación?

A pesar de la cantidad ingente de investigación que se ha realizado con el objeto de responder a esa pregunta (véase López-Bascuas, 1994a; 1994b, para una revisión) aún no se dispone de una respuesta precisa que satisfaga a la mayoría de los investigadores. No obstante, este trabajo ha señalado una nueva línea de pensamiento que podría resultar especialmente fructífera. En concreto, la idea de que procesos estrictamente auditivos pudieran aplicarse de modo selectivo al tratamiento de señales de habla. Rosner y Pickering (1994) incorporan explícitamente esta idea en un modelo de percepción de vocales que pretende ser computacional. López-Bascuas *et al.* (1995) y Fahey y López-Bascuas (1994) han proporcionado cierta evidencia experimental que apoya esta idea. Parece, pues, que este nuevo enfoque podría ayudar en el intento de resolver viejos problemas.

REFERENCIAS

- Allen, G.D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics*, 3, 91-102.
- Allen, G. D. (1981). Suprasegmental constraints on segmental representation: research involving speech production. In T. Myers, J. Laver & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland.
- Bell, A. & Hooper, J.B. (Eds.) (1978). *Syllables and segments*. Amsterdam: North-Holland.
- Cole, R., Rudnicky, A., Zue, V. & Reddy, D.R. (1980). Speech as patterns on paper. In R. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J.: LEA.
- Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, 19, 141-177.
- Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385-400.
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Chomsky, N. & Miller, G.A. (1963). Introduction to formal analysis of natural languages. In R.D. Luce, R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2. New York: Wiley.
- Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. New York: Harper-Row.
- Delattre, P.C., Liberman, A.M. & Cooper, F.S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 27, 769-773.
- Delgutte, B. & Kiang, N.Y.S. (1984). Speech coding in the auditory nerve I: Vowel-like sounds. *Journal of the Acoustical Society of America*, 75, 866-878.
- Elman, J.L. (1989). Connectionist approaches to acoustic/phonetic processing. In W. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.
- Fahey, R.P. & López-Bascuas, L.E. (1994). The perception of vowel height in Castilian Spanish: Effects of varying F1-F0 Bark distance. *Journal of Acoustical Society of America*, 96, 3283 (A).
- Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. *Logos*, 5, 3-17.
- Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT press.
- Fodor, J.A. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Frauenfelder, U., Seguí, J. & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, 16, 77-91.
- Fromkin, V. (1980). *Errors in linguistic performance*. New York: Academic Press.
- García-Albea, J.E. (1991). La mente como máquina simbólica. *Revista de Occidente*, 119, 47-60.
- García-Albea, J.E. (1992). *La capacidad humana del lenguaje: un ejemplo de discontinuidad evolutiva*. Conferencia pronunciada en el CSIC.
- Glasberg, B.R. & Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47, 103-138.
- Greenberg, J.H. (1963). *Universals in language*. Cambridge, MA: MIT Press.
- Greenberg, J.H. (1966). Synchronic and diachronic universals in phonology. *Language*, 42, 508-517.
- Halle, M. (1964). On the bases of phonology. In J.A. Fodor & J.J. Katz (Eds.), *The structure of language*. Englewood Cliffs, N.J.: Prentice-Hall.
- Halle, M. & Stevens, K.H. (1979). Some reflections on the theoretical basis of phonetics. In B. Lindblom & S. Öhrman (Eds.), *Frontiers of speech communication research*. London: Academic Press.
- Hockett, C.F. (1958). *A course in modern linguistics*. New York: MacMillan.
- Hoemeke, K.A. & Dichl, R.L. (1994). Perception of vowel height: The role of F1-F0 distance. *Journal of the Acoustical Society of America*, 96, 661-674.
- Hombert, J.M., Ohala, J.J. & Ewan, W.G. (1979). Phonetic explanations for the development of tones. *Language*, 55, 37-58.
- Jakobson, R., Fant, G. & Halle, M. (1951). *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.
- Jakobson, R. & Halle, M. (1956). *Fundamentals of language*. The Hague: Mouton.
- Jamieson, D.G. (1987). Studies of possible psychoacoustic factors underlying speech perception. In M.E.H. Schouten (Ed.), *The psychophysics of speech perception*. Dordrecht: Martinus Nijhoff Publishers.
- Jamieson, D.G. & Slawinska, E.B. (1983). Sensitivity to rate-of-change of frequency transition. *Journal of the Acoustical Society of America*, 74, Suplemento 1, S67.
- Jamieson, D.G. & Slawinska, E.B. (1984). The discriminability of transition duration: Effects of the amplitude and duration of the following steady state. *Journal of the Acoustical Society of America*, 76, Suplemento 1, S29.
- Kaye, J. (1989). *Phonology: A cognitive view*. Hillsdale, N.J.: LEA.

- Kenstowicz, M. & Kisseberth, C. (1979). *Generative phonology*. New York: Academic Press.
- Kewley-Port, D. (1983). Time varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73, 322-335.
- Klatt, D.H. (1978). SCRIBER and LAFS: Two new approaches to speech analysis. In W.A. Lea (Ed.), *Trends in speech recognition*. New York: Prentice-Hall.
- Klatt, D.H. (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279-312.
- Klatt, D.H. (1982). Speech processing strategies based on auditory models. In R. Carlson & B. Ganstrom (Eds.), *The representation of speech in the peripheral auditory system*. New York: Elsevier Biomedical Press.
- Klatt, D.H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.
- Lahiri, A. & Marslen-Wilson, W.D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 254-294.
- Lahiri, A. & Marslen-Wilson, W.D. (1992). Lexical processing and phonological representation. In G.J. Docherty & D.R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, segment, prosody*. Cambridge, England: Cambridge University Press.
- Lieberman, A.M. (1982). On finding that speech is special. *American Psychologist*, 37, 148-167.
- Lieberman, A.M., Delattre, P.C., Cooper, F.S. & Gerstman, L.H. (1954). The role of consonant-vowel transitions in the perception of the stops and nasal consonants. *Psychological Monographs*, 68, 1-13.
- Lieberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, 74, 431-461.
- Lieberman, A.M. & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. Lebowitz & H.L. Teuber (Eds.), *Handbook of sensory physiology*, Vol. XIII. Heidelberg: Springer (pp. 143-178).
- Lindau, M. (1978). Vowel features. *Language*, 54, 541-563.
- Lindblom, B. (1986). On the origin and purpose of discreteness and invariance in sound patterns. In J.S. Perkell & D.H. Klatt (Eds.), *Invariance and variability in speech processes*. Hillsdale, N.J.: LEA (pp. 493-510).
- López-Bascuas, L.E. (1994a). Procesamiento auditivo general y procesamiento específico en la percepción del habla (I): efectos derivados de la asignación de fronteras perceptivas. *Estudios de Psicología*, 52, 37-53.
- López-Bascuas, L.E. (1994b). Procesamiento auditivo general y procesamiento específico en la percepción del habla (II): efectos de integración perceptiva. *Estudios de Psicología*, 52, 55-66.
- López-Bascuas, L.E., Fahey, R.P., Hernández, M.V. & Gómez, A. (1995). Acoustic analysis of castilian front vowels: measurements of F1-F0 Bark distance. *Actas del II Simposium de Psicolinguística*, p.57.
- Marr, D. (1982). *Vision*. New York: Freeman.
- Marslen-Wilson, W.D. (1993). Issues of process and representations in lexical access. In G.T.M. Altmann & R. Schillock (Eds.), *Cognitive models of language processes: The second Sperlonga meeting*. Hove, England: Lawrence Erlbaum.
- Marslen-Wilson, W.D. & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, Phonemes and Features. *Psychological Review*, 101, 653-675.
- McClelland, J.L. & Elman, J.L. (1986). The trace model of speech perception. *Cognitive Psychology*, 18, 1-86.
- McClelland, J.L., Rumelhart, D.H. & Hinton, G.E. (1986). The appeal of parallel distributed processing. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing*, vol. 1. Cambridge, MA: MIT Press/Bradford Books.
- McQueen, J.M. (1991). The influence of the lexicon on phonetic categorisation: Stimulus quality and word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 433-443.
- Miller, J.L. (1981). Effects of speaking rate on segmental distinctions. In P.D. Eimas & J.L. Miller (Eds.), *Perspectives on the studies of speech*. Hillsdale, N.J.: LEA.
- Miller, J.L. & Jusczyk, P.W. (1989). Seeking the neurobiological basis of speech perception. *Cognition*, 33, 117-137.
- Moll, K., Zimmerman, G.N. & Smith, A. (1976). The study of speech production as a human neuromotor system. In M. Sawashima & F. Cooper (Eds.), *Dynamic aspects of speech production*. Tokyo: Univ. of Tokyo Press.
- Nearey, T.M. (1981). The psychological reality of phonological representations: experimental evidence. In T. Myers, J. Laver & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland.
- Nooteboom, S.G. (1981). Speech rate and segmental perception or the role of words in phoneme identification. In T. Myers, J. Laver & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland.
- Norris, D. (1992). Connectionism: A new breed of bottom-up model. In R. Reilly & N. Sharkey (Eds.), *Connectionist approaches to natural language processing*. Hove, UK: Erlbaum.

- Norris, D. (1993). Bottom-up connectionist models of interaction. In R. Shillock & G. Altmann (Eds.), *Cognitive models of speech processing*: Sperlonga II. Hove, UK: Erlbaum.
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52, 189-234.
- Ohala, J.J. (1976). A model of speech aerodynamics. *Report of the phonology laboratory*, 1 (93-107). University of California at Berkeley.
- Ohala, J.J. (1981). Articulatory constraints on the cognitive representation of speech. In T. Myers, J. Laver & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland.
- Pisoni, D.B. & Luce, P.A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, 25, 21-52.
- Potter, R., Kopp, G. & Kopp, H. (1966). *Visible Speech*. New York: Dover Publications.
- Pylshyn, Z.W. (1984). *Computation and cognition: toward a foundation for cognitive science*. Cambridge, MA: MIT press.
- Quilis, A. y Fernández, J.A. (1992). *Curso de fonética y fonología españolas*. Madrid: Consejo Superior de Investigaciones Científicas.
- Radeau, M., Morais, J. & Seguí, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology*, 21, 1297-1311.
- Read, C. (1971). Preschool children's knowledge of English phonology. *Harvard Educational Review*, 41, 1-34.
- Repp, B.H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92, 81-110.
- Rosner, B.S. & Pickering, J.B. (1994). *Vowel perception and production*. Oxford: Oxford University Press.
- Rumelhart, D.H. & McClelland, J.L. (1986). PDP models and general issues in cognitive science. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing*, vol. 1. Cambridge, MA: MIT Press/Bradford Books.
- Sachs, M.B. & Young, E.D. (1979). Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rates. *Journal of the Acoustical Society of America*, 66, 470-479.
- Sapir, E. (1963). The psychological reality of phonemes. In D. Mandelbaum (Ed.), *Selected writings of Edward Sapir*. Berkeley: University of California Press.
- Searle, C.L., Jacobson, J.Z. & Rayment, S.G. (1979). Stop consonant discrimination based on human audition. *Journal of the Acoustical Society of America*, 65, 799-809.
- Sebastián-Gallés, N., Dupoux, E., Seguí, J. & Mehler, J. (1992). Contrasting syllabic effects in catalan and spanish. *Journal of Memory and Language*, 31, 18-32.
- Sinex, D.G. & Geisler, C.D. (1983). Response of the auditory nerve fibers to consonant-vowel syllables. *Journal of the Acoustical Society of America*, 73, 602-615.
- Stevens, K.N. (1972a). Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. *Proceedings of the seventh international congress of phonetic sciences* (206-232). The Hague: Mouton.
- Stevens, K.N. (1972b). The quantal nature of speech: Evidence from articulatory-acoustic data. In P.B. Denes & E.E. David Jr. (Eds.), *Human communication: A unified view*. New York: McGraw-Hill.
- Stevens, K.H. (1981). Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics and psychoacoustics. In T. Myers, J. Laver & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland.
- Stevens, K.H. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.
- Summerfield, Q. (1975). Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. *Speech Perception series 2*, 4 (61-72). University of Belfast.
- Valian, V. (1979). The wherefores and therefores of the competence-performance distinction. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic Studies presented to Merrill Garrett*. Hillsdale, N.J.: LEA.
- Wang, M.D. & Bilger, R.C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America*, 54, 1248-1266.
- Watson, C.S. & Foyle, D.C. (1985). Central factors in the discrimination and identification of complex sounds. *Journal of the Acoustical Society of America*, 78, 375-380.
- Yeni-Komshian, G.H. & Soli, S.D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 70, 966-975.
- Young, E.D. & Sachs, M.B. (1981). Processing of speech in the peripheral auditory system. In T. Myers, J. Laver & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland.
- Zwicker, B., Terhardt, E. & Paulus, E. (1979). Automatic speech recognition using psychoacoustic models. *Journal of the Acoustical Society of America*, 65, 487-498.

