

**LA FONÉTICA PERCEPTIVA:
TRASCENDENCIA LINGÜÍSTICA DE MECANISMOS
NEUROPSICOFISIOLÓGICOS**

VICTORIA MARRERO AGUIAR
Universidad Nacional de Educación a Distancia
vmarrero@flog.uned.es

RESUMEN

En las páginas que siguen revisamos las bases principales desde las que puede abordarse el estudio de la fonética desde una perspectiva perceptiva. Tras situarla en un contexto lingüístico, comenzamos analizando sus principales condicionantes: la neurofisiología auditiva y su funcionamiento psicoacústico, así como las relaciones entre percepción y producción. El núcleo de nuestra presentación es un resumen de los principales resultados obtenidos en el estudio perceptivo de los niveles subsegmental y segmental del sistema fónico español.

Palabras clave: *fonética, percepción, español.*

ABSTRACT

In the following pages we review the main bases from which the study of phonetics can be approached from a perceptual point of view. After placing it in a linguistic context, we discuss their main constraints: the neurophysiological patterns and its psychoacoustical functioning, as well as the relationships between perception and production. The core of our presentation is a summary of major findings in the perceptual study of subsegmental and segmental levels in the Spanish phonic system.

Keywords: *phonetics, perception, Spanish.*

1. INTRODUCCIÓN

Hace 30 años, cuando se creó el Laboratorio de Fonética de la Universidad de Barcelona, en esta disciplina la perspectiva articulatoria llevaba siglos de ventaja a la recién nacida fonética acústica¹. Precisamente la instalación de diversos laboratorios en universidades españolas, precedidos por el que Tomás Navarro Tomás dirigió en el Consejo Superior de Investigaciones Científicas (Zamora Vicente 2001), permitió el desarrollo de la descripción acústica de los sonidos lingüísticos, primero en el ámbito investigador y pronto también en el docente.

Esta presentación concierne al tercer punto de vista desde el que puede estudiarse el sistema fónico de una lengua: el perceptivo. Tal como ocurrió con su dos *hermanas mayores*, necesita ganarse un espacio diferenciado a partir de la interdisciplinariedad: sólo partiendo de algo de fisiología, algo de neurología, algo de psicoacústica, y bastante de fonética articulatoria y acústica se podrán explicar los mecanismos que subyacen a la descodificación de los sonidos. Pero la Fonética Perceptiva no es neurología, ni anatomía, ni fisiología, ni psicoacústica del habla.

A diferencia de esas ciencias próximas, como en toda disciplina lingüística, el principio organizador de la Fonética Perceptiva es el sistema de elementos que permite la comunicación oral. Por lo tanto, habría dos límites que marcarían las fronteras en esta perspectiva de la fonética:

1. La estructura sistemática que presentan todos los fenómenos lingüísticos, incluidos los del nivel fónico.
2. La finalidad comunicativa hacia la que se orienta toda esa compleja estructura².

¹ Si consideramos el inicio de la Fonética los estudios de Panini (cuya descripción del sánscrito se basaba en lugares de articulación, «esfuerzo interno» equivalente en cierta medida al modo de articulación, y «esfuerzo externo», Deshpande 2001), la delantera de la orientación articulatoria sería milenaria.

² Mientras en el primer punto creemos que habría consenso, el segundo forma parte de un debate en el que no entraremos aquí, que durante muchos años opuso a las corrientes funcionales de la lingüística, más centradas en el uso comunicativo (punto 2), y a las corrientes formales, más interesadas en el análisis de la competencia interna (punto 1). Hoy

Así pues, de la misma manera que antes de abordar la descripción articulatoria de las lenguas se hace necesario conocer la anatomía de los órganos articulatorios, y del mismo modo que una descripción acústica no se puede abordar sin cierta familiarización previa con conceptos básicos de física de ondas, la Fonética Perceptiva requiere una descripción inicial de la anatomofisiología del sistema auditivo y del funcionamiento de la corteza auditiva cuyo nivel de profundidad vendrá determinado por las necesidades de un lingüista: habremos de entender las diferencias en la descodificación de vocales y consonantes, por ejemplo, pero no tanto los mecanismos de ubicación espacial de la fuente sonora...

Igualmente, necesitaremos comprender el funcionamiento psicoacústico del sistema auditivo, y conocer los límites de la audición. Pero sólo en la medida en que afecta a los hechos de lenguaje. Una muestra: nos interesa el umbral absoluto de la audición porque juega un papel esencial como límite para la percepción del habla, pero también, y más concretamente, por las diferencias entre las medidas de umbral obtenidas con tonos puros y las obtenidas con material lingüístico (Moore 2006). Tenemos que explicar fenómenos como el enmascaramiento, o las bandas críticas, porque son los que determinan la representación perceptiva de las vocales, o la codificación neuropsicofisiológica de la barra de explosión de las oclusivas...

El eje central de una Fonética Perceptiva, tal como la concebimos, estaría dedicado, por lo tanto, a la descodificación de los sonidos lingüísticos, desde sus componentes menores (los rasgos distintivos, desde un punto de vista estructural; las Diferencias Mínimas Perceptibles, desde el punto de vista funcional), hasta las de mayor alcance, los suprasegmentos, pasando por los segmentos y sus combinaciones en la sílaba³.

Paralelamente, y así como desde las perspectivas articulatoria o acústica se enlaza con modelos teóricos de funcionamiento del nivel fónico de las lenguas, la Fonética Perceptiva trasciende la pura descripción (anatomoneurofisiológica o psicoacústica), para aportar soluciones a las formulaciones teóricas, no solo fonológicas, sino también psicolingüísticas (en cuyo marco se desarrollan los modelos de percepción del habla).

en día creemos que la mayor parte de las propuestas se encaminan a visiones abarcadoras, como la que ofrece, en el nivel fónico, la Teoría de la Optimidad (Hayes, 1996).

³ Intentar aprehender los fenómenos de la descodificación en toda su amplitud nos llevaría mucho más allá del sonido, hasta los niveles de la palabra, la oración, el texto y el discurso, que se alejan demasiado del ámbito en el que ahora nos estamos situando

Por último, posiblemente sería también oportuna una reflexión sobre las aplicaciones que esta perspectiva de la fonética tiene en campos como la enseñanza de segundas lenguas y los déficits perceptivos.

En esta presentación nos centraremos en ese foco principal que mencionábamos: la descodificación de los sonidos lingüísticos. Comenzaremos analizando brevemente algunos de sus condicionantes generales. A continuación entraremos en el nivel subsegmental, y segmental; el suprasegmental desborda los límites preestablecidos para esta presentación, aunque su contribución es esencial para el proceso de la descodificación lingüística.

2. LOS CONDICIONANTES GENERALES DE LA DESCODIFICACIÓN LINGÜÍSTICA

2.1. El instrumento: su funcionamiento

El sistema auditivo está constituido por el oído, el nervio y las vías auditivas y la corteza cerebral, concretamente el área auditiva. A lo largo de los pocos centímetros que van desde la oreja hasta la corteza auditiva se va procesando la señal de entrada, una onda acústica, para transformarla primero en movimiento mecánico del líquido endococlear, a continuación en descargas eléctricas, y acabar conformando patrones complejos de estimulación. Para conseguirlo, será necesaria la intervención de múltiples mecanismos; algunos ni siquiera los mencionaremos. Sólo vamos a esquematizar los efectos que producen sobre la onda sonora de entrada, y sus principales funciones para la descodificación lingüística.

El oído se divide para su estudio, como sabemos, en tres partes: oído externo, medio e interno. El primero empieza en el pabellón auditivo (oreja o *pinna*) y termina al final del conducto o canal auditivo (*meatus*). Su función, además de ayudar a la localización del sonido, es de ligera amplificación de las intensidades en un rango de frecuencias medio-alto. Concretamente el conducto auditivo resuena a unos 3.000 Hz, incrementando esa zona de frecuencias unos 12-15 dB⁴.

El oído medio comienza en la membrana timpánica y concluye en la ventana oval, que es la puerta de entrada a la cóclea. Se comporta como un potente amplificador

⁴ Para realizar este cálculo se toma como base el funcionamiento de un tubo de 28 mm. (la longitud media del canal auditivo humano en el adulto) cerrado por un extremo.

con control de ganancia: si no existiera, sólo un 0,1% de la energía que llega al tímpano saldría del oído interno; pero a la vez, varios mecanismos protegen a este delicado órgano de los traumas que le podrían producir intensidades sonoras excesivas.

En el tímpano encontramos la primera y más básica separación de frecuencias del oído y el primer mecanismo de control de intensidad. En cuanto a la primera, si la onda es grave, el tímpano vibra como un todo, pero en altas frecuencias, distintas partes de la membrana responden a distintas frecuencias. Respecto al segundo, si la onda que llega es demasiado intensa el tímpano se tensa, vibrando menos. Pero los verdaderos protagonistas de esa enorme capacidad amplificadora del oído medio son los tres elementos que constituyen la *cadena de huesecillos*: martillo, yunque y estribo, los más pequeños del cuerpo, insertos en el más duro (el hueso temporal), para unir el tímpano y la cóclea. Y precisamente la diferencia de tamaño entre esos dos extremos -la membrana timpánica y la ventana oval, donde el estribo presiona la cóclea- explica buena parte de esa acción amplificadora: la proporción de 17/1 (la primera mide unos 55 mm², y la segunda apenas 3,2 mm²) incrementa la presión unas 35 veces. Pero, además, nos encontramos con el efecto palanca obtenido gracias a la diferencia de tamaño entre el martillo y el yunque: otro 30% de incremento en la energía de entrada.

Paradójicamente, unida a esta capacidad amplificadora, también encontramos entre las funciones de estos huesos diminutos la contraria, la protección frente a los excesos sonoros. Para ello, contamos con el llamado *reflejo acústico*, una contracción involuntaria del músculo estapedial ante todo estímulo inferior a 1-2 KHz e intensidad superior a 85-90 dB, cuya consecuencia inmediata es sustituir el efecto palanca al que acabamos de aludir por una acción rotatoria: 20 dB menos. Ocurre -muy interesante para el apartado 2.3.- que mientras hablamos podemos desencadenar en nuestro propio oído el reflejo acústico, cerrando el «canal de entrada» comunicativo.

Sobre la selección de frecuencias, el oído medio interviene, aunque no de manera tan potente como respecto a la energía: sus músculos actúan como un filtro de paso bajo, con una atenuación aproximada de 15 dB por octava en la zona de 1.000 Hz.

Continuamos nuestro camino para llegar a la cóclea o caracol, el órgano de la audición por excelencia, y elemento principal del oído interno⁵. La cóclea es una diminuta central eléctrica. En sus escasos tres centímetros y medio se transforma el

⁵ Porque los canales semicirculares, que la acompañan, son de la mayor importancia para los estudiosos del equilibrio, pero no para los del lenguaje.

movimiento del líquido coclear en descargas que activarán las fibras nerviosas hasta llegar al cerebro. Para conseguirlo contamos con dos líquidos cuyo contacto suelta chispas, en el sentido literal, gracias a sus diferentes composiciones químicas: la *endolinfa*, presente en el interior del conducto coclear⁶, rica en sodio y pobre en potasio y la *perilinf*a, en el interior de las células ciliadas (y en el resto de la cóclea), que es justo lo contrario: rica en potasio y pobre en sodio. La presión del estribo sobre la ventana oval provoca una *ola* en el viscoso líquido coclear, la *onda viajera* o *desplazante*, que avanza desde la base hasta el ápice. Ese movimiento, diferente para cada frecuencia de entrada, produce a su vez una fricción de los cilios, que provoca el intercambio de iones de sodio y potasio entre la endolinfa y la perilinfa. Consecuencia: despolarización de la célula ciliada, por eso aumenta la actividad eléctrica de la fibra nerviosa a ella conectada: el chispazo.

En el caracol se realiza un análisis frecuencial de la onda de entrada afinadísimo. Primero, y de forma más basta, gracias a la *selectividad frecuencial*: la capacidad del sistema auditivo para responder de modo selectivo según la frecuencia del estímulo, para filtrar un estímulo frente a otros simultáneos sobre la base de su frecuencia. Nos referimos a ese patrón de la onda desplazante, diferente según la frecuencia del estímulo (los sonidos agudos presentan su punto de máxima amplitud al inicio de la cóclea, y cuanto más graves más se desplazan hacia el final): es la «mecánica coclear pasiva». A ella se une otro mecanismo (calificado como «activo»), del que son responsables las células ciliadas externas, las más abundantes (un 75% del total). La suma de ambos es una descomposición frecuencial muy afinada del estímulo de entrada (sobre todo para sonidos de frecuencias altas e intensidades medias y bajas; el resto necesita otro tipo de procesamiento, que veremos a continuación), transformado en un patrón de estimulación eléctrica, apto para recorrer el camino hacia el cerebro (Gil Loyzaga y Poch Broto 2000)

En cuanto al control de energía, el oído interno también contribuye, gracias a lo que se conoce como *comportamiento no lineal* de la cóclea. En los seres vivos, los incrementos de intensidad en la señal no se traducen en idénticos incrementos en la amplitud de vibración de la membrana basilar. Llega un momento en que, por más

⁶ La cóclea se divide en dos rampas por medio de un canal hueco, el *conducto coclear*, en cuya base se inserta el *órgano de Corti*, donde residen a su vez las *células ciliadas*: unos 15.000 receptores neurosensoriales, que se denominan así por sus *estereocilios* (“vellos”) rígidos, cuyo roce con la membrana superior del propio órgano de Corti abre y cierra los canales iónicos por donde se intercambian los componentes químicos de la endolinfa y la perilinfa, provocando los cambios de potencial eléctrico

intensidad que se aplique en la estimulación, la respuesta obtenida se estanca paulatinamente.

La acción del nervio auditivo, ya en el siguiente estadio de procesamiento, refuerza la selectividad frecuencial de la cóclea, porque sus fibras también presentan una *organización tonotópica*: las fibras responden mejor a rangos de frecuencia determinadas. Pero además está especializado en un mecanismo que también presentaban las células ciliadas: la *sincronización temporal*.

Las neuronas se caracterizan por su *tasa de disparo*, el número de activaciones que pueden mantener por unidad de tiempo. En el nervio auditivo esta tasa es de 400-500 disparos por segundo. Cuando la fibra se sincroniza temporalmente con la onda sonora el disparo se produce sólo en un punto preciso del ciclo de la onda. Pero sabemos que el habla tiene muchos componentes importantes por debajo de los 400 ciclos por segundo. La solución para codificarlos es la sincronización de varias neuronas con una sola onda, de tal modo que mientras una está en fase de reposo, otra toma el relevo. El patrón de respuesta sería la combinación de todas ellas.

Los dos principios, el frecuencial y el temporal, se combinan para codificar la señal de entrada de la manera más resistente posible.

En el nervio auditivo encontramos, además, un comportamiento no lineal, tanto ante la duración del estímulo (la fibra «se cansa» y deja de responder a partir de los primeros 15-20 ms.) como ante las frecuencias (enmascaramiento de un componente de la onda sobre otro). Ambos fenómenos permiten resaltar las diferencias entre sonidos sucesivos: explotar la dinámica en la señal de habla, sacar partido al eje temporal.

Las vías auditivas (núcleo coclear, complejo olivar superior, lemnisco lateral, colículo inferior y cuerpo geniculado medio) son un complejo sistema de haces neuronales que van y vienen, arriba y abajo, en el tronco encefálico, para acabar en la corteza auditiva. Es el conjunto de interconexiones más complicado de todo el sistema sensorial: el 70% de las vías son contralaterales, cruzan desde un oído a la corteza cerebral del lado contrario; el resto son ipsilaterales. Hay vías nerviosas que atraviesan todas las estaciones intermedias hasta llegar a la corteza, mientras que otras se saltan algunos núcleos. Esta complejidad es necesaria para compensar la inicial pobreza neurológica del sistema auditivo. Si lo comparamos con el visual, por ejemplo, partimos de un déficit del 0,01%⁷. Al llegar al cerebro, ambos siste-

⁷ La retina tiene 130 millones de receptores fotosensibles, frente a las 15.000 células ciliadas del oído; el nervio óptico está formado por un millón de fibras, el auditivo sólo contiene 40.000 - 50.000 fibras.

mas se equilibran: encontramos unos cien millones de neuronas, tanto en el cortex visual como en el auditivo. Esto es posible gracias a las vías auditivas, y su sinuoso y complejo camino hacia la corteza cerebral: en cada una de las estaciones intermedias se multiplica el número de neuronas (Handel, 1993).

Las vías auditivas mantienen en todas sus estaciones la organización tonotópica de la cóclea y del nervio auditivo: cada neurona tiene una frecuencia característica, en la cual la intensidad necesaria para activarla es menor, tiene el umbral más bajo.

Por otra parte, a medida que avanzamos hacia la corteza cerebral, las neuronas tienden a responder mejor a las partes dinámicas del habla (transiciones, movimientos de los formantes, inicios, finales: puntos de cambio espectral) y a las variaciones de la frecuencia fundamental.

Y llegamos por fin al cerebro, la cima del sistema nervioso. El área auditiva de la corteza (o córtex auditivo) se encuentra en el interior de la cisura de Silvio, perteneciente al lóbulo temporal (entre cuyas funciones destacan la audición, el aprendizaje, la memoria y las emociones). Conecta con el área de Wernicke, encargada de la descodificación de las unidades lingüísticas, una zona asociativa superior de la mayor importancia para el procesamiento del lenguaje.

La corteza auditiva está compuesta por neuronas con una alta conectividad y organizadas en columnas jerárquicas; las del hemisferio izquierdo presentan un diámetro mayor y más espacio entre ellas que las del derecho. Se suelen distinguir en ella dos zonas muy precisas: el área auditiva primaria y la secundaria.

El área auditiva primaria está organizada tonotópicamente en bandas de isofrecuencias: las frecuencias bajas activan las zonas más externas de la cisura de Silvio, y las altas zonas más profundas. En cuanto al procesamiento temporal, en esta área se encuentran dos poblaciones de neuronas:

1. Sincronizadas, responden a cambios temporales lentos (más de 20 ms); serían las encargadas de procesar el tono del sonido.
2. No sincronizadas, responden a cambios temporales más rápidos (10-20 ms), lo que les permite detectar las fronteras entre sonidos (Gil-Loyzaga 2005: 34-35)

Por último, ciertas unidades concretas se encargarían de integrar estos dos tipos de información mediante una respuesta dual.

El área auditiva secundaria, en cambio, no tiene distribución tonotópica. Se encarga de la localización del sonido, y del procesamiento de patrones auditivos complejos; posiblemente tiene relación con la memoria auditiva.

Alrededor de ambas áreas se encuentra la región periférica, donde la audición se integra con el resto de los sistemas sensoriales, por lo que se denomina también área de integración sensorial polimodal. Aquí es donde la información visual y la auditiva se ponen en relación.

En esencia, el papel de las regiones corticales consistiría en realizar las funciones superiores de integración del mensaje oral. El procesamiento de los eventos auditivos (definidos como elementos relativamente concurrentes, en este caso neuronales) formaría un patrón de excitación de codificación de la información con carácter distintivo, mucho menos variable que el espectro acústico (Greenberg 1996: 391). Las neuronas actúan aquí por comparación con patrones aprendidos.

En la tabla 1 se esquematiza la información presentada en este apartado. Necesariamente hemos dejado sin mencionar muchos aspectos, algunos importantes, como la codificación de la intensidad (centrándonos solo en los mecanismos de atenuación o amplificación), y otros muchos de detalle.

		Control de intensidad		Selección de frecuencias	Análisis de duraciones
		Amplificación	Atenuación		
Oído ext.	Pabellón auditivo	Rango medio-alto			
	Conducto auditivo	3 KHz → + 15 dB			
O. medio	Timpano		Tensión timpánica	Graves → vibr. total Agudas → vibr. parcial	
	Huesecillos	Diferencia de área → *35 Efecto palanca → + 30%	Reflejo acústico: -20dB	Filtro paso bajo (1 KHz)	
O. int.	Cóclea		Respuesta no lineal a la intensidad	Organización tonotópica Mecánica activa	
Nervio auditivo				Organización tonotópica Sincronización temporal Enmascaramiento	Adaptación
Vías auditivas				Organización tonotópica Sincronización temporal	Respuesta a elementos dinámicos
Corteza auditiva				Organización tonotópica Integración de patrones	Neuronas sincronizadas y no sincronizadas Respuesta a elementos dinámicos Integración de patrones

Tabla 1. El sistema auditivo y la descodificación de los sonidos lingüísticos.

2.2. El instrumento: sus límites

El funcionamiento del sistema auditivo, que hemos presentado muy sintéticamente, determina los límites de nuestra capacidad decodificadora. Sus fronteras vienen determinadas, como siempre, por la intersección de las tres cualidades físicas principales del sonido: intensidad, duración y frecuencia.

1. *Frecuencia.* Podemos oír sonidos desde 20-30 Hz hasta 20.000, pero somos especialmente sensibles a los que se sitúan entre 2.500 y 5.000 Hz.

2. *Intensidad.* La capacidad para percibir cambios de intensidad en el oído humano normal crece logarítmicamente más de cien veces desde el umbral de audición (0 dB⁸ en la mayoría, aunque algunas personas alcancen los -10 dB) hasta el umbral del dolor (110-120 dB). La proporción entre el sonido más intenso previo al dolor, y el más débil que podemos percibir es de 1.000.000.000.000/1, o lo que es lo mismo, $10^{12}/1$.

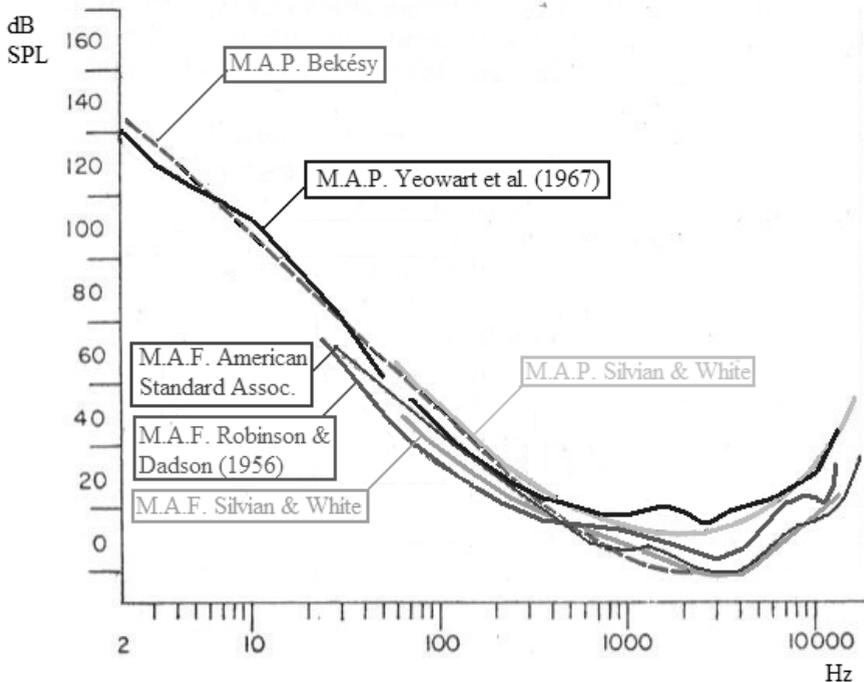
3. *Duración.* A este respecto sólo tenemos un límite, y es el inferior: el sonido lingüístico más breve perceptible puede oscilar entre 10 y 40 ms. No hay techo temporal en percepción auditiva (otra cosa son las restricciones debidas a la capacidad de atención, que no vienen al caso). Pero nuestra mayor sensibilidad natural, según varios estudios, aparece en el rango que va de 40 a 60 ms⁹.

Tenemos que considerar, además, lo que ocurre cuando se entrecruzan estas tres categorías, y especialmente las dos primeras, porque no todas las frecuencias requieren la misma intensidad para ser percibidas. Muy al contrario, existe una zona, la que va de 500 a 8000 Hz, mucho más sensible que el resto, donde apenas

⁸ Ocurre que las medidas de intensidad utilizadas en el estudio de la percepción humana están fijadas de forma convencional, y precisamente en relación con nuestra capacidad auditiva. El nivel 0 dB se sitúa justamente en el umbral absoluto de audición para un determinado tono (1 kHz para dB SPL –*Sound Pressure Level*-, diversas frecuencias para los dBHL –*Hearing Level*-).

⁹ Esa 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 (López Bascuas 1997: 12)

es necesario subir de 0 dB para percibir un sonido; por encima y por debajo necesitaremos incrementar la intensidad significativamente para alcanzar los mismos resultados, como puede verse en las curvas de audibilidad de la figura 1.



Adaptado de Guirao (1980) y López Bascuas (1998)

Figura 1. *Curvas de audibilidad.* (M.A.P. = Minimal Audible Pressure ; se obtienen mediante auriculares. M.A.F.= Minimal Audible Field ; se obtienen en campo libre).

En la intersección entre los topes en frecuencias (infrasonidos / ultrasonidos) y los de intensidades (umbral de audición / umbral del dolor), obtenemos lo que se ha denominado el *campo de audición*, o *área de respuesta auditiva*, muy utilizado en entornos audiológicos (figura 2). En él se observa de forma muy gráfica cómo dos intensidades muy diferentes pueden producirnos la misma sensación subjetiva de

«fuerza»: un tono de 30 Hz necesitará 65 dB para producirnos la misma sensación perceptiva que otro de 1000 Hz, con apenas 0-2 dB. Y también, como *grosso modo*, el área conversacional se sitúa en pleno centro del campo de audición.

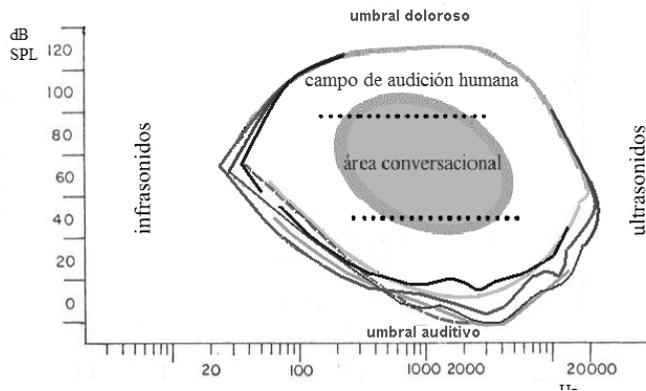


Figura 2. Campo de audición.

En contextos psicofísicos aparece más a menudo otra forma de plasmar esa misma información: las *curvas de isofonía*, o *contornos equisonoros* (figura 3), una representación más detallada, que nos explica qué ocurre con cada frecuencia dentro del campo de audición (figura 4).

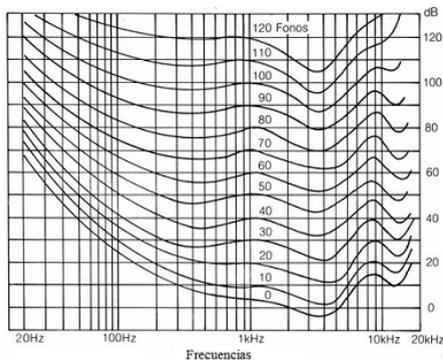


Figura 3. Curvas de isofonía.

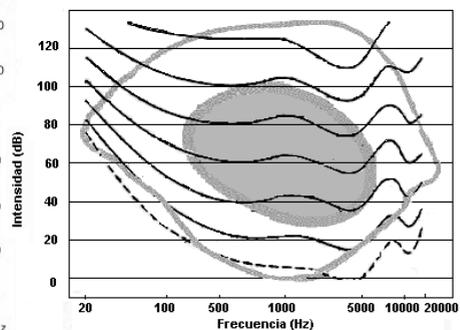


Figura 4. Curvas isofonía y campo aud.

Desde un punto de vista lingüístico, lo interesante es relacionar esta información con la procedente de la acústica del habla, y más concretamente con el espectro acústico del español, como el que encontramos en la figura 5:

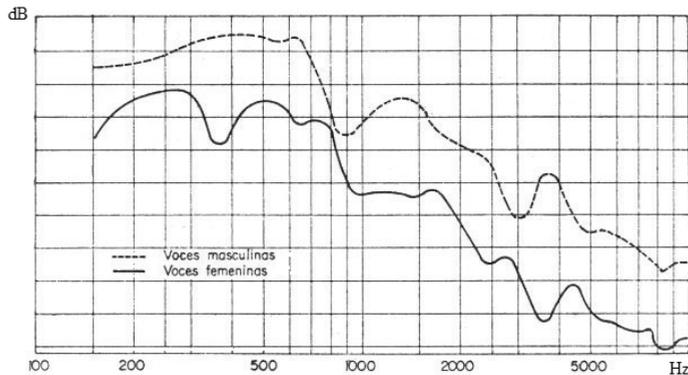


Figura 5. *Espectro acústico del español; promedio de energía. Adaptado de Guirao (1980: 201).*

Los componentes espectrales del habla de frecuencias más bajas aparecen entre 150 y 800 Hz, coincidiendo con un descenso en las curvas de isofonía (índice de mayor sensibilidad), al que se suma una mayor intensidad entre 1.500 y 5.000 Hz: la zona de mayor receptividad auditiva¹⁰.

2.2.1. Enmascaramiento y bandas críticas

En general, cuando la presencia de un estímulo interfiere con la percepción de otro, decimos que el primero está *enmascarando* al segundo. La definición oficial (de la *American Standards Association*, 1960) de este fenómeno es *el proceso por el cual el umbral de audibilidad de un sonido aumenta debido a la presencia de otro sonido (máscara)*. Se trata, posiblemente, del mecanismo más estudiado en psicofísica auditiva. Desde sus primeras manifestaciones científicas, puso de

¹⁰ El pico que los contornos equisonoros muestran en frecuencias altas (8 KHz) no tiene un correlato claro en el promedio de energía del español, según los datos de Guirao (1980). No obstante, resulta inevitable recordar la importancia de esta zona de frecuencias para la discriminación de /s/, cuyo peso morfológico en español es de primera magnitud

manifiesto ciertas peculiaridades, contradicciones de las leyes generales de la física, que constituirían características específicas de la capacidad analítica del sistema auditivo, consecuencia del funcionamiento de la membrana basilar, con repercusiones en las células ciliadas y en las fibras del nervio auditivo y las vías cocleares.

Nos referimos a respuestas no lineales, a irregularidades como las que quedan de manifiesto en estas curvas: todo lo que queda por debajo de la línea base sería la señal enmascarada, que no puede oírse.

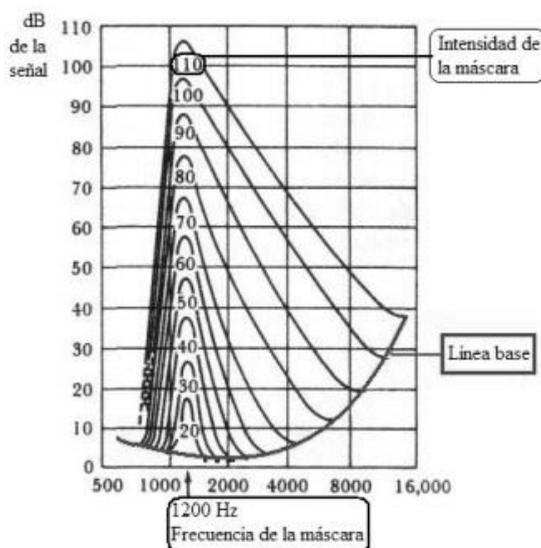


Figura 6. *Curvas de enmascaramiento.*

Como vemos, el efecto enmascarador es más fuerte en la frecuencia que coincide con la de la máscara (donde aparece el pico de la gráfica en la figura 6).

Por otra parte, cuanto mayor es la intensidad de la máscara, más amplia es el área enmascarada. Pero tal efecto se incrementa más hacia las frecuencias agudas (a la derecha) que hacia las graves: es lo que se conoce como *extensión hacia arriba* o *enmascaramiento remoto* (Corso 1979: 489). Por tanto los sonidos graves

enmascaran a los agudos, y no a la inversa. Esto afecta a la percepción de los sonidos del lenguaje, y se ha revelado como especialmente relevante en el procesamiento de las transiciones

Los estudios sobre el ancho de banda del enmascaramiento en el oído permitieron descubrir el sistema de *bandas críticas* que caracteriza la audición. Se denomina así a cada una de las fronteras o límites -24 en total- en que el sistema auditivo agrupa las frecuencias para su procesamiento. Funcionan como filtros de paso de banda, con un ancho mayor para las frecuencias graves, hasta 500 Hz (lo cual permite el procesamiento de los armónicos del F0, por ejemplo), y menor por encima de 1 KHz (para el procesamiento de formantes). *La percepción auditiva se basa en el análisis en bandas críticas del oído interno. De hecho, el analizador auditivo periférico puede ser concebido como un banco de filtros de paso de banda, numerados de $x=1$ a $x=24$, con anchos de banda de alrededor de 100 Hz por debajo de 500 Hz y anchos de banda de $1/6$ del centro de la frecuencia por encima de 1 kHz* (Schroeder, Atal y Hall 1979: 220).

Así, los estímulos que coincidan en el interior de una banda crítica interactúan perceptivamente (el más intenso predomina), pero los que pertenezcan a bandas diferentes se analizan de modo independiente.

2.3. Las relaciones con el exterior

El exterior del sistema perceptivo humano es el mundo del sonido, y más concretamente, por lo que a nuestros intereses concierne, el sistema de sonidos del lenguaje, y su origen, los mecanismos de la producción del mensaje oral.

En el apartado anterior mostrábamos cómo el área conversacional se sitúa en el corazón del campo de audibilidad. En este, veremos un poco más detalladamente las relaciones entre percepción y producción. En primer lugar abordaremos el origen –filogenético y ontogenético- de esa relación necesaria entre dos sistemas con restricciones muy distintas, pero condenados a entenderse.

Desde el punto de vista evolutivo, el sistema auditivo precede en varios millones de años al fonador¹¹. Éste último, muy especializado en nuestra especie, nos permite un repertorio segmental y suprasegmental mucho más amplio que el de los

¹¹ La historia del aparato auditivo en el ser humano se remonta al periodo cretácico (hace más de 6,5 millones de años), aunque algunos de sus rasgos básicos (como la sincronización en fase y la organización tonotópica) son muy anteriores, del paleozoico, 225 millones de años atrás. En cambio nuestro aparato fonador actual sólo se remonta a unos 150.000 años.

primates, nuestros parientes más cercanos. En cambio, el sistema auditivo humano apenas difiere del que disfruta el resto de los mamíferos¹². Por lo tanto, parece muy probable que las innovaciones del aparato fonador vinieran determinadas, en buena medida, por las restricciones previas que el sistema auditivo imponía para constituirse en canal de transmisión hacia el cerebro de toda información oral: la producción se habría amoldado a las condiciones impuestas por la percepción. (Greenberg y Ainsworth 2004)

Pero a su vez, el sistema auditivo se ha tenido que ir adaptando a las condiciones ambientales, del entorno, como un instrumento lo suficientemente sensible para discriminar sonidos relevantes biológicamente en las condiciones habituales del mundo natural. Y ninguno más relevante para la especie humana que el habla.

Las relaciones entre los sonidos que percibimos y los que emitimos, por otra parte, se ponen de manifiesto muy claramente en el desarrollo lingüístico del bebé. Una extensa línea de investigación sobre capacidades perceptivas tempranas demuestra que los niños nacen con cierta predisposición innata para la discriminación de estímulos sonoros del tipo que encontramos en todas las lenguas naturales. Pero a lo largo de los seis primeros meses de vida esa habilidad natural se va especializando, reduciendo, a los sonidos más frecuentes en el entorno del recién nacido (Jusczyk 1997). Simultáneamente, el balbuceo, sus primeras producciones orales (más allá de los ruidos biológicos), reproduce igualmente ciertas características de esos sonidos del entorno, a los que se va ajustando mejor progresiva pero rápidamente antes de llegar al segundo año¹³.

De hecho, se considera probable que la aparición de categorías lingüísticas abstractas (sílabas, fonemas, unidades con significado) en el cerebro del niño esté motivada, entre otras razones, por la necesidad de compatibilizar las

¹² En el nivel de la corteza auditiva, evidentemente, las diferencias son enormes.

¹³ En la historia de los estudios sobre adquisición del lenguaje ha habido una larga polémica a este respecto, iniciada con las propuestas de Jakobson sobre la discontinuidad entre el balbuceo y las primeras palabras del niño. Para el lingüista praguense, se trataría de dos etapas sin relación entre sí, la primera obedecería a leyes universales, comunes a todas las lenguas, mientras que la segunda ya sufriría las restricciones particulares de cada lengua. Esta hipótesis se ha ido rebatiendo con el tiempo, y hoy en día parece probado que el balbuceo presenta características propias en cada lengua, y que se trata de un entrenamiento imprescindible para emisiones posteriores con significado, sin rupturas entre la etapa prelingüística y la lingüística, precisamente creando lazos entre producción y percepción.

representaciones obtenidas de estas dos procedencias: la propia producción y la percepción del habla. Para conseguir ese nivel de abstracción se generalizarían los rasgos más relevantes y comunes a ambos sistemas.

La retroalimentación, el *feedback*, entre lo auditivo y lo vocal continúa a lo largo de toda la vida. Una milésima de segundo después de emitir cualquier cadena fónica oímos nuestra propia voz, tanto a través de los huesos del cráneo como del aire. Inmediatamente, ese mensaje es procesado en la corteza auditiva (Eliades y Wang 2004) y se inicia un proceso de auto-monitorización¹⁴. Incluso en presencia de importantes déficits auditivos, que impiden el desencadenamiento de este proceso, bastan unas semanas de entrenamiento con un laringógrafo no sólo para controlar las variaciones tonales de la voz propia, sino también para aumentar la capacidad para diferenciar entonaciones ajenas (Fourcin 1979). El papel de la retroalimentación vocal sobre la percepción es algo más controvertido. Sin embargo, se ha convertido en bandera de una de las más conocidas teorías sobre percepción del habla: la *Teoría Motora* defiende que para descodificar cualquier sonido necesitamos crear previamente una representación mental de sus movimientos articulatorios correspondientes (Liberman y Mattingly 1985)

En todo caso, percepción y producción están sujetas a presiones contrapuestas: por la ley del mínimo esfuerzo, la tendencia del aparato fonoarticulatorio es la mayor relajación posible, la neutralización máxima, la producción de todos los sonidos en las zonas más próximas. Sin embargo, la descodificación requiere todo lo contrario: la máxima distancia entre unos elementos y otros, enfatizar sus diferencias, incrementar sus distancias perceptivas. El resultado de esta contienda entre máxima pronunciabilidad y máxima discriminabilidad es el compromiso que suponen los sonidos del lenguaje: un *contraste suficiente* (según terminología de Lindblom 1992)¹⁵, que ha permitido la supervivencia de los mejores elementos desde ambas perspectivas: sonidos fáciles de distinguir y poco costosos de

¹⁴ Cuando la retroalimentación se retrasa artificialmente aparecen alteraciones transitorias en la producción oral: reducción de la tasa de habla; aumento de la intensidad y del F0; incluso una especie de tartamudez provocada, con repetición de sílabas o consonantes, omisiones de sonidos o de partes de la palabra, sustituciones, etc. Este descubrimiento de Lee y Fairbanks en los años 50 permitió la aplicación del feedback auditivo retrasado en el tratamiento de la disfemia (tartamudez) y otras patologías.

¹⁵ Este autor parte de esta relación entre lo que percibimos y lo que producimos para construir su *Teoría de la Variabilidad Adaptativa*, conocida también como *Hipótesis de Hipo e Hiperarticulación (H&H)*.

producir. Es la selección natural que reduce a 35 o 40 elementos la enorme gama de vocalizaciones que nuestro aparato fonador nos permite.

El deseo, la necesidad de ser entendidos nos empuja a armonizar ambos sistemas. Tal como decía Jakobson hace 45 años, *we speak to be heard in order to be understood*.

3. ELEMENTOS MÍNIMOS

Hasta este momento hemos abordado las principales restricciones externas que condicionan el sistema perceptivo, considerado un instrumento a través del cual tiene lugar la descodificación de los mensajes lingüísticos. A partir de ahora entraremos en el principal objetivo de esta presentación: la estructura interna de los mecanismos de descodificación. Comenzaremos por sus componentes menores, en el nivel subsegmental: los rasgos distintivos, si tenemos en cuenta la estructura del sistema – perspectiva lingüística- y las Diferencias Mínimas Perceptibles (DMP), si nos situamos en una perspectiva funcional –psicolingüística-.

3.1. Los rasgos distintivos

Como ocurre ante cualquier fenómeno inexplorado, las investigaciones sobre percepción del habla han partido de lo conocido, para relacionarlo con lo nuevo. Por eso se han buscado correspondencias entre los mecanismos de descodificación y uno de los elementos más ampliamente asentados en los estudios fónicos: los rasgos distintivos.

En ocasiones, se han tomado como referencia las tradicionales clasificaciones articulatorias: modo de articulación, lugar de articulación y sonoridad/sordez. Entre los resultados obtenidos por este tipo de experimentos (realizados fundamentalmente mediante manipulación del habla, suprimiendo o enmascarando ciertas bandas de frecuencia), destacamos los siguientes (O’Shaughnessy 1990):

1. El modo de articulación es el rasgo articulatorio más resistente. Las diferencias entre oclusión, fricación, etc., se han mantenido en circunstancias adversas con más contundencia que las que oponen sonidos sordos a sonoros, o labiales a dentales, palatales o velares. Sus claves perceptivas parecen residir en frecuencias inferiores a 1000 Hz

2. El lugar de articulación, en cambio, presentó la mayor fragilidad. Las frecuencias más relacionadas con su percepción son las situadas por encima de 1000 Hz, especialmente la zona de los segundos formantes. Se ve afectado por la superposición de ruido de banda ancha (por su incidencia en F2 y F3), y por la reverberación (se experimentó con un eco de 800 ms.)
3. La sonoridad depende, como era de esperar, de la estructura de los armónicos, más fuerte en frecuencias muy bajas, pero que se mantiene incluso hasta los 3000 Hz. Por esa razón, una manipulación en frecuencias inferiores a 1KHz provoca confusiones entre pares de sonidos sordo/sonoro, pero el ruido de banda ancha apenas afecta. Retomaremos estos tres puntos en 4.2.

Otros estudios se han centrado en la percepción de los rasgos distintivos acústicos. Nos interesa destacar entre ellos un trabajo de hace dos décadas, más conocido entre audiólogos que entre fonetistas, por su carácter aplicado a las pruebas de logaudiometría. Lo firma una lingüista india, Vasanta Duggirala, en colaboración con el equipo formado por tres eminentes audiólogos (Pavlovic, Studebaker y Sherbecoe). En él se establece una relación entre los rasgos acústicos y una determinada frecuencia crítica (*crossover frequency*), determinante para su percepción¹⁶ (Duggirala et al. 1998):

Nasalidad:	472 Hz	Densidad:	1618 Hz
Sonoridad:	758 Hz	Continuidad:	1800 Hz
Gravedad:	1290 Hz	Estridencia:	2521 Hz.

Estos datos han sido confirmados en experimentos sobre el español con oídos patológicos: existe una relación entre cada rasgo distintivo y una banda de frecuencia determinada: a mayor pérdida auditiva en esa frecuencia, peores

¹⁶ En la metodología empleada se tomó como estímulo un test construido con pares mínimos opuestos por un rasgo distintivo, sobre el cual se aplicó el Índice de Articulación (*Articulation Index*, una medida establecida en 1947 por French y Steinberg para determinar la cantidad de información lingüística preservada en la audición residual de las personas con pérdida auditiva). Se determinó así la importancia relativa de las diferentes partes del espectro auditivo en la discriminación de los rasgos distintivos acústicos.

resultados en el rasgo distintivo correspondiente. Cuando un sujeto tiene afectada su capacidad para percibir sonidos agudos, por ejemplo (el tipo de pérdida auditiva más frecuente en lesiones neurosensoriales) las oposiciones peor identificadas son 'continuo/interrumpido' y 'estridente/mate' (Marrero, Santos y Cárdenas 1993; Marrero y Martín 2001).

Como vemos, los esfuerzos de los investigadores se han centrado en la búsqueda de claves acústicas responsables de la percepción de unos rasgos u otros. En esta presentación, necesariamente breve, sólo hemos ofrecido unas pinceladas sobre el tipo de resultados obtenidos.

3.2. Las Diferencias Mínimas Perceptibles (DMP)

Los estudios de percepción definen las Diferencias Mínimas Perceptibles o Diferencia Apenas Perceptible (*Just Noticeable Difference, JND*) como la menor diferencia que puede ser detectada entre dos estímulos: la unidad mínima en las escalas de sensación (Tudela 1989: 33). Aunque constituyen una fértil área de estudio sobre distintos sentidos (especialmente visión y audición), y distintos tipos de estímulos, nosotros nos limitaremos a sus aplicaciones a los sonidos del lenguaje: lo que se conoce como psicofísica del habla o *psicofonética*¹⁷:

3.2.1. Frecuencias

En condiciones experimentales (vocales sintéticas en estado estable) nos basta un 3-5% de variación en los valores centrales de los formantes para detectar el cambio. Pero si el estímulo es un poco más natural (sílabas abiertas) necesitaremos incrementar esas variaciones entre un 9 y un 14% para que sean perceptibles (O' Shaughnessy 1990: 157-159). Cuando manipulamos el ancho de banda formántico habremos de saltar al 20-40% y alterar también la intensidad para obtener un resultado perceptivo; al variar concomitantemente su duración afectaremos de forma especial a los fonemas nasales.

3.2.2. Frecuencia Fundamental.

Nuestra sensibilidad perceptiva es extrema en lo que respecta al F_0 : los estudios pioneros de Flanagan en 1957 con vocales abiertas –en las que el F_1 no interfiere

¹⁷ Terminología de T'Hart, Collier y Cohen, 1990.

con el F_0 podemos notar variaciones del 0,2-0,5%, inferiores a un Herzio, especialmente si se trata de una subida tonal, más que de un descenso. Sin embargo, en habla más natural los cambios del F_0 inferiores a 5 Hz, y 50 msg, son promediados durante la percepción. *Diferencias en el F_0 de menos de un cuarto de octava es improbable que sean útiles en la comunicación lingüística* (O'Shaughnessy, 1990: 158).

En un interesante trabajo sobre umbrales tonales en español peninsular, desde el Laboratori de Fonètica de la U.B. se llega a la conclusión de que, la diferencia mínima perceptible para los cambios frecuenciales en entonación es de un semitono, pero en su uso lingüístico *ceñirse al mínimo perceptible no debe ser funcionalmente rentable*; necesitamos un poco más de redundancia, para evitar que cualquier distorsión de la señal haga irrecuperable el mensaje. Por eso se introduce el concepto de *umbral funcional*, estableciéndolo en 1,5 semitonos, *la unidad mínima potencialmente relevante desde el punto de vista prosódico* (Pamies et al. 2001).

En definitiva, los cambios de F_0 lingüísticamente relevantes son mucho más marcados que los límites de la sensibilidad psicofísica: el habla está construida por medio de claves robustas, a las que se dota de valor significativo, con un nivel de abstracción elevado. Para ello, además, contamos con la información proveniente de la interrelación entre todos los elementos presentes en la onda sonora del habla: no sólo frecuencias, sino también su interacción con intensidades y duraciones.

3.2.3. Intensidades

Según los estudios iniciales de Flanagan, anteriormente mencionados, el umbral diferencial de intensidad promedio para formantes en estímulos lingüísticos es de 2 dB. Sin embargo, esa media se distribuye de forma desigual en el espectro: el primer formante, que es el más prominente perceptivamente (*saliente*) requiere tan sólo 1,5 dB, mientras que el F2 debe duplicar esa cifra para provocar un cambio perceptivo; las frecuencias interformánticas necesitan alcanzar incluso 13 dB.

3.2.4. Duraciones

Las DMP temporales en sonidos del lenguaje han sido objeto de bastantes estudios, que nos proporcionan resultados muy variados: desde 3 msg hasta 36. Tal disparidad se debe, en buena parte, a diferencias metodológicas entre los

experimentos realizados: diferentes tipos de estímulos, distintas tareas, lenguas tipológicamente distintas, nivel de entrenamiento de los jueces variado, etc.

Con las precauciones necesarias, y teniendo en cuenta esa variabilidad, podemos presumir que las mínimas diferencias perceptibles en lo temporal para los sonidos del lenguaje son bastante pequeñas, aunque varían de unos rasgos a otros: 10 ms de señal bastan para percibir, en vocales aisladas, un cambio en el lugar de articulación; la percepción de la sonoridad puede requerir entre 20 y 30 ms. Para las transiciones necesitamos un 25-30% de los valores de referencia¹⁸. Y somos más sensibles a diferencias de duración en vocales que en consonantes.

Precisamente las duraciones vocálicas, y en concreto las del español, han sido objeto de estudio por Pàmies y Fernández Planas (2006)¹⁹. Su resultado proporciona un umbral perceptivo diferencial del 35,9%: el incremento más pequeño que podemos percibir sobre la vocal más breve sería de unos 36 ms. Esta cifra es más alta que la de muchos estudios anteriores, pero se considera *un umbral representativo de la media* sobre duraciones habituales en habla natural, y en una lengua sin diferencias fonológicas de cantidad (Pàmies y Fernández, 2006: 509).

Con las vocales concluimos este apartado sobre elementos mínimos (tanto del análisis fonético como de las escalas perceptivas), y llegamos al siguiente: las menores unidades lingüísticas, los fonemas.

4. SEGMENTOS

4.1. La percepción de las vocales

Casi un axioma en fonética acústica, originado en los laboratorios Haskins durante la década de los 50, ha sido considerar el primer y segundo formantes como las

¹⁸ Eso hace pensar que quizá sólo manejemos dos categorías de transiciones: larga / breve

¹⁹ En su experimento se utilizó habla natural (la sílaba [mám]), cuidadosamente manipulada (para uniformar la intensidad, el tono, los formantes y las transiciones), de modo que la única diferencia entre los estímulos fuera de orden temporal (100 ms de duración base, incrementada en escalones del 10%, 20%, 25%, 33%, 40%, 50%, 66%, 75% y 100%), presentada en condiciones naturales (altavoces en un aula grande) a cien jueces no entrenados

claves acústicas definitivas para la descripción y clasificación de las vocales. ¿Bastan F1 y F2 para el análisis perceptivo de los sonidos vocálicos? La influencia del F3 para vocales adelantadas, en sistemas complejos, como el del inglés, se demostró casi desde entonces²⁰.

Pero incluso considerando el tercer formante, los experimentos basados exclusivamente en estas claves arrojan porcentajes de identificación muy bajos: en vocales naturales aisladas del inglés apenas supera el 58% (según Strange y sus colaboradores, en 1974); en contexto, la cifra ascendía hasta el 83% (Verbrugge y coautores, ese mismo año, citados ambos por Landercy y Renard, 1977). Para las vocales francesas los resultados eran algo mejores: el 80% en vocales naturales aisladas (Landercy y Renard, 1977). El español presenta, sin embargo, datos excelentes, según Borzone de Manrique (1979): 97% de identificación correcta en vocales naturales aisladas, y del 99% en contexto²¹.

En todo caso, los resultados del inglés convencieron a la mayoría de los investigadores de que un modelo de percepción vocálica que utilizara simplemente los primeros formantes sería inadecuado, especialmente para hacer frente a las tareas de normalización ante las tres fuentes de variabilidad: individual, temporal y contextual. De este modo, en los años 80, nacen los «modelos de normalización vocálica» (Disner, 1980), que tienen en cuenta información acústica complementaria procedente del F₀ y de medidas relativas (distancia entre F₂ y F₃, entre F₁ y F₀).

La frecuencia fundamental, pues, se ha revelado como un elemento muy importante en la percepción de las vocales. Halberstom y Raphael, (2004) la consideran más relevante que el F3 para normalizar las diferencias entre locutores²². Traunmüller (1987), llega incluso a defender que información sobre la

²⁰ Según Landercy y Renard (1977), en un estudio que incluía las vocales nasalizadas, éstos son los índices más importantes para cada timbre: F₁ y F₂: [a y u o æ ø ɔ]; F₁ y F₃: [e Ē y i]; F₂ y F₃: [a i].

²¹ La autora atribuye esta excelente respuesta a dos factores: las diferencias acústicas entre el sistema vocálico del español y el del inglés (menor número de unidades, y como consecuencia, mayor separación entre ellas); y el modo de percepción, que sería categorial en español y continuo en inglés.

²² Utilizaron estímulos susurrados (sin F₀) y filtrados (sin F₃) de nueve vocales emitidas por 15 locutores (5 hombres, 5 mujeres y 5 niños). Según sus resultados, se acertaron más vocales filtradas (86%) que susurradas (77%).

fase en el F_0 sería suficiente para diferenciar la cualidad de la vocal si el tono es suficientemente bajo (voz masculina). Las diferencias entre la representación perceptiva y la representación acústica de una vocal pueden estar en la base de esta revalorización del F_0 : en los espectros auditivos (cfr. Figura 7) la frecuencia fundamental presenta la misma resolución que los formantes, y ocupa un área perceptiva notablemente mayor que la obtenida en una representación acústica (lo cual también ocurre con los dos primeros formantes)²³.

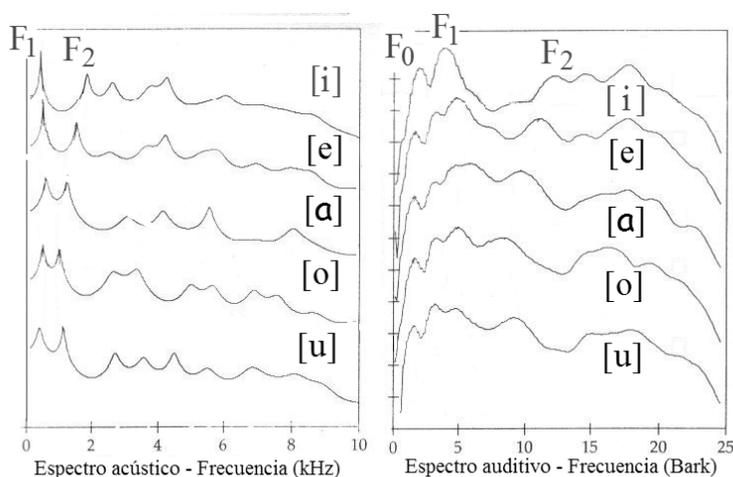


Figura 7. *Espectros acústicos y auditivos de las vocales. Adaptado de Johnson (1997:116).*

¿Percibimos de la misma manera todas las vocales? ¿Utilizamos las mismas claves perceptivas para las anteriores y las posteriores, las abiertas y las cerradas? Las diferencias de timbre abrieron una nueva vía de investigación sobre percepción vocálica, especialmente de la mano de Ludmilla Chistovich al frente del Instituto

²³ Las propiedades del sistema auditivo son las responsables de estas diferencias: las bandas críticas, suficientemente estrechas en bajas frecuencias, actúan como un mecanismo de integración frecuencial, fusionando los primeros armónicos del F_0 y resolviéndolos como un pico separado. La respuesta no-lineal de la periferia auditiva, por su parte, refuerza la importancia de las regiones del F_1 y el F_2 .

Pavlov, en Moscú. Propone la existencia de un «centro de gravedad» perceptivo determinante para ciertas vocales: mientras las anteriores dependerían del F1 y el F2, en las posteriores la cercanía entre ambos (inferior a una distancia crítica que se estableció en 3-3.5 barks) conllevaría su integración en una sola clave: el centro de gravedad perceptivo, situado entre los dos formantes y característico de cada vocal (Chistovich, Sheikin y Lublinskaja 1979)

Pero ocurre también que en las vocales cerradas el F1 está muy cerca del F0, provocando igualmente la integración perceptiva, dominada por el punto más cercano entre ambos elementos: el armónico del F0 más cercano al pico del F₁. La existencia de este mecanismo en español ha sido estudiada por Fahey y López Bascuas (1994). Según sus resultados, la frontera entre vocal alta / baja se establecería cuando la distancia entre F0 y F1 estuviera por debajo o por encima de 2.5 barks (por tanto, la establecida por Chistovich y sus colegas no constituiría un fenómeno universal, derivado de discontinuidades del sistema auditivo, sino que variaría de unas lenguas a otras).

¿Son igualmente perceptibles todas las vocales? ¿Y lo son en todas las lenguas? Los estudios de campos de dispersión nos permiten dar respuestas a algunas de estas preguntas. Menos comunes que los campos de dispersión acústicos, los perceptivos se realizan presentando a un conjunto de jueces una serie de estímulos vocálicos (naturales o artificiales) y pidiéndoles que los identifiquen como una de las vocales de su sistema.

En la carta de formantes de la figura 8 hemos representado los principales resultados de los estudios sobre campos de dispersión perceptivos realizados para el español peninsular²⁴: Romero (1989) -rectángulos rellenos- y Fernández Planas (1993) -trazo discontinuo-.

A diferencia de los habituales campos de dispersión acústicos, los perceptivos muestran unas áreas considerablemente más amplias²⁵, excepto en el caso de [a], el

²⁴ Romero presenta también un estudio con jueces bilingües catalán-español. Los campos de dispersión perceptivos de las vocales catalanas han sido estudiados por Puig y Freixa, 1990, y Fernández Planas y Carrera Sabaté 2004. Para el español de Chile contamos también con el trabajo de León Valdés 1998.

²⁵ Las diferencias evidentes entre los resultados obtenidos por ambos estudios pueden deberse a la distinta metodología empleada, entre otros factores: Romero utiliza voz natural (15 vocales cardinales más 3 inglesas); Fernández Planas emplea síntesis de voz: 371 estímulos, con un F1 que iba de 109 a 1119 Hz, y un F2 de 500 a 2450.

elemento con menor variabilidad perceptiva. Las vocales cerradas, en cambio, presentan un campo de dispersión bastante amplio, pero ubicado en los extremos del rectángulo, lejos de las realizaciones vecinas. [o] reduce drásticamente su campo a las realizaciones más abiertas (especialmente en los resultados con estímulos naturales), interponiendo la mayor de las distancias con [u]. En cambio, [e] se permite cierto margen de variación en este rasgo, radicado en el primer formante.

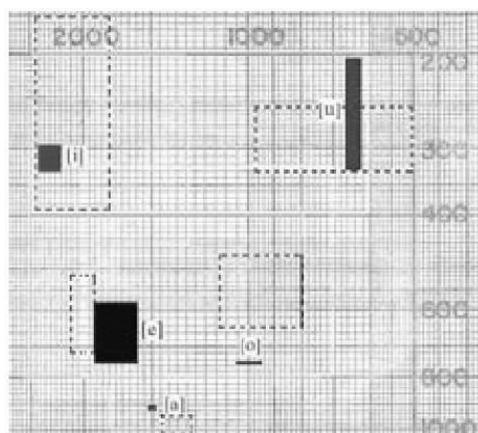


Figura 8. Campos de dispersión perceptivos de las vocales españolas. Romero (1989) –rectángulos llenos- y Fernández Planas (1993) –trazo discontinuo-.

En cuanto a los mecanismos neurofisiológicos encargados de codificar la información que se enviará al cerebro, lo más probable es que recurramos a la combinación de la selectividad frecuencial (sólo se excitan las fibras que coinciden en frecuencia con el estímulo; apto para intensidades bajas) y la sincronización temporal en fase (cada fibra se activa, independientemente de su frecuencia de respuesta, en determinado punto del ciclo del estímulo; apto durante los primeros 30 msg), dada la riqueza armónica de los sonidos vocálicos.

En definitiva, la percepción vocálica, y la general, se muestra como un mecanismo adaptativo, que necesita hacer frente a la necesidad de hacer frente a las fuentes de variabilidad de la señal de habla (sonidos extranjeros, sonidos artificiales,

diferentes locutores, etc.), que han de ser normalizadas; de ahí la amplitud de los campos de dispersión perceptivos. También la tarea que ha de resolver el sujeto influye en el tipo de representación vocálica (Espinoza-Varas, 1987): para el etiquetado fonético bastaría una más basta, sensible sólo a cambios en varias bandas críticas. Los juicios sobre buena cualidad vocálica requerirían una representación más detallada, en la cual las variaciones del F_2 tendrían más repercusiones que para el simple etiquetado.

4.2. La percepción de las consonantes

La percepción de las consonantes suele ser más difícil que la de las vocales. Encontramos razones para ello en sus características acústicas como estímulo: menor riqueza en componentes armónicos; menor intensidad y, a menudo, inferior duración. Pero también, probablemente, en el uso de diferentes tipos de clave para la identificación de unas y otras: mientras para las vocales las claves espectrales tienen un papel primordial (sea por selectividad frecuencial, sea por sincronización en fase), como hemos visto, en la identificación de las consonantes podrían ser prioritarias las claves temporales (cambios rápidos en el estímulo, como sucede en las transiciones, en el VOT²⁶, en las barras de explosión, etc.).

A continuación revisaremos muy rápidamente las principales conclusiones obtenidas del estudio de percepción consonántica, especialmente sobre el español. Éstos han prestado mucha atención, desde sus inicios, a las oclusivas ([p, t, k], o [b, d, g]). Aunque su espectro es, en principio, lo más simple que uno pueda imaginar como sonido lingüístico (silencio + explosión), su identificación resultó, desde el primer momento, una caja de sorpresas: en la década de los 50, Delattre, Liberman y Cooper sacaron a la luz unos experimentos donde mostraban que ni el silencio ni la explosión permitían distinguir entre sí /p, t, k/, sólo los movimientos de los formantes vocálicos circundantes. Investigaciones posteriores, con estímulos más semejantes a la voz natural que los rudimentarios métodos de síntesis vocal empleados por aquel entonces en los Laboratorios Haskins, pusieron de manifiesto que la situación era mucho más compleja: las características específicas de cada

²⁶ *Voice Onset Time*, en español *Tiempo de Inicio de la Sonoridad (TIS)*, es el periodo que transcurre entre la salida turbulenta del aire fonador, al deshacerse la oclusión bucal (barra de explosión), y la vibración de las cuerdas vocales para la emisión de la vocal siguiente (barra de sonoridad). Este tiempo puede ser positivo (primero aparece la barra de explosión, y luego la de sonoridad), lo cual es característico de las consonantes sordas; o negativo (comienza la sonoridad cuando todavía no se ha deshecho la oclusión bucal, y el aire continúa retenido): es propio de las consonantes sonoras.

transición determinan su importancia perceptiva; cuando son muy lentas, y alejan la barra de explosión del centro del formante, resultan esenciales para percibir la consonante; por el contrario, cuando son breves, como la explosión está más cerca del pico espectral correspondiente al núcleo vocálico, es la explosión la que contribuye de manera determinante a la identificación de la consonante. Por otra parte, su peso perceptivo también varía según los lugares de articulación de las consonantes en cuestión, y las vocales que las rodean.

Los parámetros responsables del modo de articulación han resultado, por lo general, menos complejos que los de lugares de articulación. La percepción del rasgo ‘interrupido’, generado por una oclusión o cierre en la salida del aire fonador, parece venir determinado por ese periodo de silencio característico en este orden consonántico: *para la discriminación del rasgo oclusivo es suficiente [...] el silencio, con un tiempo no menor a 70-80 ms.* (Martínez Cedrán, 1991: 128). En cambio, para la percepción de las fricativas, parece determinante la distribución espectral del ruido, especialmente en regiones de bajas frecuencias, inferiores a 6 KHz (Feijóo, Fernández y Balsa, 1999a).

La percepción de los lugares de articulación, en cambio, resulta más compleja. Según varios autores, los oyentes realizamos esa tarea durante los primeros 20 o 30 ms. del sonido, pero mucho de lo que ocurre durante esos instantes sigue siendo una incógnita: cuáles son las claves perceptivas más prominentes (transiciones, VOT, barra de explosión, murmullo nasal, formantes en las aproximantes o líquidas...), para qué sonidos y en qué contextos. Un conjunto de trabajos han ofrecido respuestas, no siempre coincidentes²⁷, a una parte de esas preguntas:

1. ¿Existen lugares de articulación más perceptibles que otros? Establecer escalas de perceptibilidad no ha resultado tarea simple. Si nos centramos sólo en las oclusivas, Mújica, Santos y Herraiz (1990), obtienen, en la percepción de transición + consonante, la escala /t/>/p/>/k/; Elejabeitia,

²⁷ Es importante señalar que se trata de trabajos con importantes diferencias metodológicas entre sí: unos utilizan palabras (Moreno Llaneza 1990), otros sílabas (Mújica, Santos y Herraiz 1990), otros logotomas (Elejabeitia, Iríbar y Pagola 1995); en unos casos se intercambiaron claves acústicas de unas oclusivas con otras (Moreno Llaneza 1990, Elejabeitia, Iríbar y Pagola 1995), en otros se eliminaron elementos (Mújica, Santos y Herraiz 1990, Martínez Cedrán 1991, Feijoo, Fernández y Balsa 1996). Las diferencias en los resultados pueden estar en parte relacionadas con tales variaciones.

Iríbar y Pagola 1995 presentan el resultado opuesto: /k/>/p/>/t/, en el cual coinciden con Feijoo, Fernández y Balsa (1999^a).

2. ¿Cuál es el papel de las transiciones vocálicas, frente a las claves acústicas propias de cada consonantes? En el caso de las oclusivas, se ha comparado su papel con el del VOT o la barra de explosión. Llisterri, (1987) afirma que la duración de la barra de explosión y el VOT son los únicos parámetros que permiten la discriminación entre los tres lugares de articulación, mediante una combinación de mecanismos temporales (duración, especialmente significativa para la velar sorda), y frecuenciales (distribución espectral del ruido explosivo, más estable para la dental y bilabial). Coincidiendo con las escalas anteriores, Martínez Celdrán (1991) sostiene igualmente que *la velar es la que más necesita la barra de explosión para ser identificada, luego la dental, y en último término la bilabial* (Martínez Celdrán, 1991: 121). En el caso de las nasales, Albalá (1992) defiende la prioridad de las transiciones sobre el murmullo nasal, aunque estudios previos (Massone (1988), Recasens (1983) sobre las implosivas del catalán) destacaban la necesidad de combinar ambos, especialmente en las nasales no palatales. En /l/ parecen predominar también las transiciones del F2 que la frecuencia del formante consonántico en su momento estable (Fernández, Garrido y de la Mota, 1989).
3. ¿Cuál de las dos transiciones predomina en la percepción de las consonantes? En lo que a oclusivas se refiere, Elejabeitia, Iríbar y Pagola (1995) consideran preponderante la primera²⁸, a diferencia de lo defendido cinco años antes por Moreno Llanceza 1990.

En definitiva, como vemos, la percepción de los lugares de articulación consonánticos parece el resultado de la interacción compleja de varios elementos, cuyo funcionamiento posiblemente sólo se explique a la luz amplia de las necesidades comunicativas de los hablantes: *“la variación sistemática dependiente del contexto es un factor que puede ser utilizado beneficiosamente, para incrementar la cantidad de información fonética que puede ser extraída de la señal de habla”* (Suomi, en 1987, citado por Feijoo, Fernández y Balsa, 1999a: 17).

²⁸ Primera transición, la que va desde el centro de la vocal anterior hacia el silencio de la consonante oclusiva. La T2 o segunda transición iría desde la barra de explosión de la oclusiva hasta el centro de la vocal.

La percepción de la sonoridad, o de las diferencias entre /p, t, k/ y /b, d, g/, por último, ha suscitado el mayor de los debates en las investigaciones sobre consonantes oclusivas. Cuatro claves acústicas se disputan la primacía: a) la barra de sonoridad (un formante de baja frecuencia originado por la vibración glotal); b) la duración del silencio (consecuencia de la interrupción en la salida del aire fonador); c) el VOT; y d) las transiciones o frecuencia de inicio del formante vocálico.

En primer lugar, es necesario tener en cuenta una condición establecida desde muy pronto, en un trabajo de Williams publicado en los *J.A.S.A.* en 1976: en español al menos, los estímulos con habla real provocan unos resultados diferentes de los que se obtienen con voz sintética. El VOT negativo, establecido por Abramson y Lisker 1972 como propio de las oclusivas sonoras del español, es una clave suficiente para la sonoridad en vocales sintéticas, pero en voz natural no es imprescindible, porque “*other properties must supply positive voicing information in its absence*” (Williams, 1976: s41). Aunque a menudo esta salvedad no se ha tenido en cuenta en los diseños experimentales, un trabajo mucho más reciente de Hernán Emilio Pérez incide también en el mismo punto: el manipular tanto la barra de sonoridad (pegando en [p, t, k] las de [b, d, g], y suprimiendo las de [b, d, g]) como las duraciones consonánticas (intercambiando las de los pares sordos con sus correlatos sonoros y viceversa) no afecta a la identificación de los estímulos de habla natural (superior al 95%), pero sí a los de voz resintetizada (la primera manipulación afecta al timbre de /k/, percibida como /t/; la segunda provoca la confusión de /b/ por /p/ y /d/ por /t/).

Volviendo a las cuatro claves para la percepción de la sonoridad,

1. La barra de sonoridad, considerada como índice esencial en los primeros estudios acústicos sobre los sonidos del español (Navarro Tomás, Alarcos, Harris, Quilis), fue puesta en duda por Martínez Celdrán (1985) mediante experimentos de filtrado, que demostraban la persistencia en la identificación de [b, d, g] tras suprimir los primeros 500 Hz de la señal. Sin embargo, Elejabeitia, Gálvez, Iribar, Morales, Rivera y Vélez (1995) siguen defendiendo que “*la barra de sonoridad parece ser el índice más potente, seguido de los tipos de VOT y, por último, la duración*” (Elejabeitia et al. 1995: 170), aunque con matices, ya que ninguno de los tres puede funcionar como índice único. También Feijóo, Fernández y Balsa (1998), en un experimento con voz natural de numerosos locutores (18, hombres y mujeres) detectan una caída del 95% al 54% en la

identificación de /b, d, g/ cuando se suprimió la barra de sonoridad en los estímulos (resultó especialmente afectada la voz femenina).

2. El VOT negativo de las oclusivas sonoras del español (comienza la sonoridad antes de la salida turbulenta del aire retenido en el cierre bucal) ha sido objeto de numerosos estudios perceptivos, sobre todo de carácter interlingüístico²⁹, sobre los que no podemos detenernos. Sólo un dato: con voz sintética, bastaría reducir este tiempo a 38 ms. para provocar el percepto de sonoridad (Martínez Celdrán 1991).
3. Reducir la duración del silencio consonántico, provocaría el mismo efecto: de nuevo según Martínez Celdrán (1991) con síntesis de voz, si es inferior a 70 ms. induce a la percepción de la consonante sonora en lugar de la sorda. Jaime Soto Barba (1994), en un ingenioso experimento de pares mínimos p/b con estímulos naturales en voz cuchicheada (94% de aciertos), llega a la conclusión de que lo relevante no es la duración de la consonante en sí (a veces las sonoras eran más largas que las sordas), sino la relación entre ésta y la duración total de la sílaba, siempre mayor para la sorda que para la sonora.
4. La influencia del contexto, por último, y más concretamente la frecuencia del primer formante en la vocal contigua, se mostró como un factor determinante en la percepción de la sonoridad o sordez en estímulos naturales manipulados (/bit bIt bid bId/ del inglés canadiense) presentados a oyentes mexicanos sin conocimientos del inglés: las vocales con F1 más bajo se interpretaban como acompañadas de /t/, y las de F1 más alto, como con /d/.

Los efectos de la coarticulación, las diferencias que surgen en función del contexto vocálico que rodea a cada consonante son otro de los aspectos destacados una y otra vez en los estudios sobre percepción consonántica. Las vocales circundantes influyen decisivamente en la perceptibilidad de las consonantes oclusivas, especialmente en [k], que si aparece rodeada de vocales velares incrementa

²⁹ Incluso inter-especie, podríamos decir: Sinnott y McArdle 2000 comparan la percepción de oyentes de habla inglesa, española, y monos, ante un continuo sintético BA-PA. La frontera de los primeros se situó en +25 ms; la de los segundos, entre -15 y +25 (media de 7.6 ms.); y la de los monos en 0 ms. Las autoras concluyen que todos los grupos tienen las mismas capacidades sensoriales, y que la frontera del inglés no es un rasgo del sistema auditivo general.

enormemente sus índices de identificación ([uku]: 14% / [iki]: 0% en Martínez Celdrán (1991); [ko]: 79% / [ki]: 3% en Feijoo, Fernández y Balsa (1999a). También las aproximantes [β δ γ] sufren en español la influencia del timbre vocálico: con /a,e,o/ las claves perceptivas relevantes son diferentes a las que se encuentran con /i,u/: Franco y Gurlekian (1985). En las fricativas, la influencia coarticulatoria de la vocal ha sido objeto de sucesivos estudios por el grupo de procesado de señal de la Universidad de Santiago. Desde una defensa inicial (*the identification of the fricatives was clearly affected by the influence of the following vowel (particularly by the front/back distinction)*) Feijóo, Fernández y Balsa, 1999b: 12), a la posición contraria: *the influence of the vowel on the fricative is perceptually irrelevant* (Fernández et. al. 2000: 1350; también Fernández y Feijóo, 2001). Por lo que respecta a las nasales, del mismo modo *la identificación de estas consonantes parece también depender de la vocal adyacente* (Massone, 1988: 28). En cuanto a las líquidas, en un trabajo sobre /l/, Fernández, Garrido y de la Mota (1989), señalan que la dependencia coarticulatoria es mayor en contextos vocálicos posteriores ([o, u]) que anteriores ([i, e]).

Como sumaráisima conclusión, creemos que todos estos resultados ponen de manifiesto nuestra capacidad para utilizar de la forma más rentable posible todos los elementos acústicos a nuestra disposición, para conseguir una descodificación óptima de la señal. Si el estímulo que se nos presenta es artificial, crearemos estrategias específicas para resolver la tarea encomendada. En situaciones naturales de habla, muy presumiblemente jugaremos con cualquiera de las claves que la estructura lingüística pone a nuestra disposición, modificando su orden de prioridad en función de nuestras necesidades comunicativas específicas.

5. REFERENCIAS BIBLIOGRÁFICAS

- ABRAMSON, A. S. y L. LISKER (1972): «Perception of Voice Timing in Spanish Stop Consonants», *The Journal of the Acoustical Society of America* (J.A.S.A.) 52, p. 175.
- ALBALÁ, M. J. (1992): «Análisis y síntesis de las consonantes nasales», *Revista de Filología Española*, LXXII, pp. 47-73.
- BORZONE DE MANRIQUE, M. I. (1979): «On the recognition of isolated Spanish vowels», en H. Hollien y P. Hollien (eds.): *Current Issues in the Phonetic Sciences*, Amsterdam, John Benjamins, pp. 677-681.

-
- CHISTOVICH, L. A.; R. L. SHEIKIN y V. V. LUBLINSKAJA (1979): «“Centres of gravity” and spectral peaks as the determinants of vowel quality», en Lindblom y Öhman (eds): *Frontiers of Speech Communication Research*, Londres, Academic Press, pp. 143-157.
- CORSO, J. F. (1979): «Audición», en B. B. Wolman (ed.): *Manual de Psicología General*, Barcelona, Martínez Roca (ed. inglesa 1973), p. 489.
- DESPANDE, M. M. (2001): «Panini and The Distinctive Features», en W. Kreidler (ed.): *Phonology*, Londres, Taylor and Francis, pp. 40-52.
- DISNER, S. F. (1980): «Evaluation of vowel normalization procedures», *The Journal of the Acoustical Society of America*, 67, (1), pp. 253-261.
- DUGGIRALA, V.; G. A. STUDEBAKER; C. V. PAVLOVIC y R. L. SHERBECOE (1988): «Frequency Importance Functions for a Feature Recognition Test Material», *The Journal of the Acoustical Society of America*, 83 (6), pp. 2372-2382.
- ELEJABEITTA, A.; A. IRÍBAR y R. M. PAGOLA (1995): «Índices de sonoridad/sordez en las oclusivas españolas», en A. Elejabeitia y A. Iribar (eds.): *Phonetica. Trabajos de fonética experimental*, Deusto, Univ. de Deusto, pp. 157-170.
- ESPIÑOZA-VARAS, B. (1987): «Involvement of the critical band in the identification, perceived distance and discrimination of vowels», en Schouten (ed): *The Psychophysics of Speech Perception*, Dordrecht, Martinus Nijhoff Pub, pp. 306-313.
- FEIJÓO, S.; S. FERNÁNDEZ y R. BALSALBA (1998): «Integration of acoustic cues in Spanish voiced stops», *Proceedings ICA/ASA meeting (Seattle)*, pp. 2933-2934.
<http://www.usc.es/proc/public.htm>
- FEIJÓO, S.; S. FERNÁNDEZ y BALSALBA, R. (1999a): «Acoustic and perceptual study of phonetic integration in Spanish voiceless stops», *Speech Communication*, 27, pp. 1-18.
- FEIJÓO, S.; S. FERNÁNDEZ y R. BALSALBA (1999b): «Influence of frequency range in the perceptual recognition of fricatives», *Proceedings of the Forum Acusticum/ASA/EAA/DEGA Congress, Acustica-Acta Acustica*, vol. 85, supl. 1, p. S475 (A).
<http://www.usc.es/proc/public.htm>
-

-
- FERNÁNDEZ PLANAS, A. M. (1993): «Estudio del campo de dispersión de las vocales castellanas», *Estudios de Fonética Experimental*, V, pp. 129-145.
- FERNÁNDEZ PLANAS, A. M. y J. CARRERA SABATÉ (2004): «Percepción de vocales medias del catalán oriental central y del catalán noroccidental por parte de oyentes catalanohablantes y castellanohablantes de Lleida», en M. Villayandre (coord): *Actas del V Congreso de Lingüística General*, León 5-8 de marzo de 2002, Vol. 1, 2004, ISBN 84-7635-574-2 , pags. 973-982.
- FERNÁNDEZ, H.; J. M. GARRIDO y C. DE LA MOTA (1989): «Modelling coarticulation in synthesized Spanish lateral consonant [l]», en T. Szende (ed.): *Proceedings of the Speech Research'89*, Budapest, Linguistic Institute of the Hungarian Academy of Sciences. Pp. 210-213.
http://liceu.uab.es/publicacions/Fernandez_Garrido_Mota_89_Coarticulation_Synthesized_Lateral.pdf
- FERNÁNDEZ, S. y S. FEIJÓO (2001) «Coarticulatory effects in perception» *Proceedings of EuroSpeech (Aalborg)*, 155-158.
<http://www.usc.es/proc/public.htm>
- FERNÁNDEZ, S.; S. FEIJÓO; R. Balsa y N. BARROS (2000): «Perceptual effects of coarticulation in fricatives», *Proceedings of ICASSP (Estambul)*, vol. 3, 1347-1350.
<http://www.usc.es/proc/public.htm>
- FOURCIN, A. J. (1979): «Auditory pattering and vocal fold vibration», en Lindblom y Öhman (eds), *Frontiers of Speech Communication Research*. Londres, Academic Press, pp. 167-175.
- FRANCO, H. y J. A. GURLEKIAN (1985): «Recognition of Spanish intervocalic consonants», *J.A.S.A.* 77 (S1), p. S27.
- GIL LOYZAGA, P. y J. POCH BROTO (2000): «Fisiología del sistema auditivo periférico», en Suárez et al. (eds.): *Tratado de otorrinolaringología y cirugía de cabeza y cuello*, Editorial Panamericana, vol. 2, pp. 762-774.
- GIL LOYZAGA, P. (2005): «Estructura y función de la corteza auditiva. Bases de la vía auditiva ascendente», en E. Salesa, E. Perelló y A. Bonavida (eds.): *Tratado de audiología*, Barcelona, Masson, pp. 23-38.

-
- GREENBERG, S. y W. A. AINSWORTH (2004): «Speech processing in the auditory system: an overview», en S. Greenberg; W. A. Ainsworth; A. N. Popper y R. R. Fay (eds.): *Speech Processing in the Auditory System*, Nueva York, Springer-Verlag, pp. 1-62.
- GREENBERG, S. (1996): «Auditory processing of speech», en Lass (ed.): *Principles of Experimental Phonetics*, Sant Louis, Mosby, pp. 362-407.
- GUIRAO, M. (1980): *Los sentidos, bases de la percepción*, Madrid, Alhambra.
- HALBERSTOM, B. y L. J. RAPHAEL (2004): «Vowel normalization: the role of fundamental frequency and upper formants», *Journal of Phonetics*, 32 (3), pp. 423-434.
- HANDEL, S. (1993): *Listening. An Introduction to the Perception of Auditory Events*, Cambridge, Massachusetts, The MIT Press.
- HAYES, B. P. (1996): «Phonetically Driven Phonology: The Role of Optimality Theory and Inductive Grounding», escrito para las actas de la Conferencia de Milwaukee *Formalism and Functionalism in Linguistics*, 1996.
<http://www.humnet.ucla.edu/linguistics/people/hayes/phonet/indgroun.pdf>
- JOHNSON, K. (1997): *Acoustic and Auditory Phonetics*, Londres, Blackwell, 2ª ed.
- JUSCZYK, P. W. (1997): *The discovery of spoken language*, Cambridge, Massachusetts, The MIT Press.
- LANDERCY, A. y R. RENARD (1977): *Éléments de phonétique*, Bruselas, Didier, 2ª ed.
- LEÓN VALDÉS, H. (1998): «Determinación del campo de dispersión auditiva de las vocales del español de Chile», *Revista de Lingüística Aplicada*, 36, pp. 87-96.
- LIBERMAN, A..M. e I. MATTINGLY (1985): «The motor theory of speech perception revised», *Cognition* 21, pp. 1-36.
- LINDBLOM, B. (1992): «Phonological units as adaptative emergents of lexical development», en C. A. Ferguson, L. Menn y C. Stoel-Gammon (Eds): *Phonological development: Models, reseach, implications*, Timonium, Md. York Press.

-
- LÓPEZ BASCUAS, L. E. (1997): «La percepción del habla: problemas y restricciones computacionales», *Anuario de Psicología*, 72, pp. 3-19.
- MARRERO, V. e Y. MARTÍN (2001): «Discriminación auditivas de los rasgos distintivos acústicos en palabras aisladas: oídos normales y patológicos», en J. Díaz García (ed): *Actas del II Congreso de Fonética Experimental*, Sevilla, Universidad de Sevilla, pp. 258-266.
- MARRERO, V.; A. SANTOS y M. R. CÁRDENAS (1993): «Feature Discrimination And Pure Tone Audiometry», en R. Aulanko e Y. Korpijaakko-Huuhka (eds): *Proceedings of the Third Congress Of The International Clinical Phonetics And Linguistics Association, 9-11 August 1993, Helsinki*, University of Helsinki, pp. 121-128.
- MARTÍNEZ CELDRÁN, E. (1991): *Fonética experimental: teoría y práctica*, Madrid, Síntesis.
- MARTÍNEZ CELDRÁN, E. (1985): «¿Hasta qué punto es importante la sonoridad en la discriminación auditiva de las obstruyentes mates del castellano?», *Estudios de Fonética Experimental*, 1, pp. 245-291.
- MASSONE, M. I. (1988): «Estudio acústico y perceptivo de las consonantes nasales y líquidas del español», *Estudios de Fonética Experimental*, III, pp. 13-34.
- MOORE, B. C. J. (2006): «Factors affecting speech intelligibility for people with cochlear hearing loss», en S. Greenberg y W. A. Ainsworth (eds): *Listening to speech: an auditory perspective*, Londres, Routledge, pp. 273-288.
- O'SHAUGHNESSY, D. (1987): *Speech Communication: Human and Machine*, Reading, Massachusetts, Addison Wesley Pub. Comp, 1990.
- PÀMIES, A. y A. M. FERNÁNDEZ PLANAS (2006): «La percepción de la duración vocálica en español», en J. Luque Durán (ed.): *Actas del V congreso andaluz de lingüística general. Homenaje a José Andrés de Molina*, Granada, pp. 501-512.
- PÀMIES, A.; A. M. FERNÁNDEZ PLANAS; E. MARTÍNEZ CELDRÁN; A. ORTEGA y M.C. AMORÓS (2001): «Umbrales tonales en español peninsular», en J. Díaz García (ed.): *Actas del II Congreso de Fonética Experimental*, Sevilla, Universidad de Sevilla, pp. 272-278.
-

-
- PUIG, J. y J. FREIXA (1990): «El camp de dispersió de les vocals catalanes des del punto de vista de la percepció», *Estudios de Fonética Experimental*, IV, pp. 123-146.
- RECASENS, D. (1983): «Place cues for nasal consonants with special reference to Catalan», *The Journal of the Acoustical Society of America*, 73 (4), pp. 1346-1353.
- ROMERO, J. (1989): «Campos de dispersión auditivos de las vocales del castellano. Percepción de las vocales», *Estudios de Fonética Experimental*, III, pp. 181-206.
- SCHROEDER, M. R; B. S. ATAL y J. L. HALL (1979): «Objective measure of certain speech signal degradation based on masking properties of human auditory perception», en Lindblom y Öhman (eds): *Frontiers of Speech Communication Research*, Londres, Academic Press, pp. 217-229.
- SINNOTT, J. M. y L. A. MCARDLE (2000): «Comparative perception of voicing contrasts», *The Journal of the Acoustical Society of America*, 107 (5), p. 2855.
- SOTO BARBA, J. (1994): «¿Los fonemas /b/ y /p/ se diferencian por la sonoridad?», *Estudios filológicos*, 29, pp. 33-37.
- T'HART, J; A. COLLIER y R. COHEN (1990): *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*, Cambridge, C.U.P.
- TRAUNMÜLLER, H. (1987): «Phase vowels», en Schouten (ed): *The Psychophysics of Speech Perception*, Dordrecht, Martinus Nijhoff Pub, pp. 377-384.
- TUDELA, P. (1989): *Psicología experimental*, Madrid, UNED.
- WILLIAMS, L. (1976): «Prevoicing as a perceptual cue for voicing in Spanish», *J.A.S.A.*, 59 (S1), p. S41.
- ZAMORA VICENTE, A. (2001): *Los orígenes de la fonética experimental en España: Exposición*, Cáceres, Fundación Biblioteca Alonso Zamora Vicente, p. 22.