

PERCEPCIÓN DEL HABLA BAJO LA PRESENCIA DE SONIDOS ENMASCARANTES. BREVE REVISIÓN DE LOS PROCESOS INVOLUCRADOS Y UNA APROXIMACIÓN AL PROCESAMIENTO AUDITIVO CENTRAL

ADRIÁN FUENTE
BRADLEY MCPHERSON
*Division of Speech and Hearing Sciences
The University of Hong Kong
Hong Kong
Hong Kong S.A.R.*

RESUMEN

La habilidad de comprender el habla bajo la presencia de varios sonidos enmascarantes es un fenómeno complejo que ya ha sido estudiado por varios autores. El fenómeno de *cocktail party* propuesto por Cherry (1953) fue el primer intento de dar a entender cómo los oyentes son capaces de utilizar claves monoaurales y binaurales cuando discriminan el habla bajo la presencia de sonidos enmascarantes. El presente artículo pretende discutir el fenómeno de percepción del habla bajo la presencia de sonidos enmascarantes, el rol de la atención y la memoria de trabajo como “importantes procesos cognitivos” que los oyentes utilizan al percibir un sonido verbal bajo la presencia de enmascaradores de información. Finalmente, en la última parte del artículo se abordan algunas pruebas de procesamiento auditivo central y las habilidades auditivas involucradas en cada prueba. Esto, tomando en consideración que las habilidades auditivas están directamente relacionadas con distintos mecanismos de la percepción del habla, especialmente bajo la presencia de sonidos enmascarantes.

ABSTRACT

The ability to understand verbal messages in the presence of multi-sounds is a complex phenomenon that has been already studied by many authors. The cocktail party phenomenon proposed by Cherry (1953) was the first attempt to understand how listeners are able to take advantage from both monaural and binaural cues. Fundamental frequency, prosody, intensity and interaural time differences are some of the cues that listeners use to segregate and group sounds. This article aims to discuss speech perception in the presence of background noise. It also aims to discuss both attention and working memory as “important cognitive processes” when listeners manage a verbal message in the presence of informational masking. The last part of the article aims to address some central auditory processing tests and the auditory abilities involved, taking into account that the auditory abilities are directly linked to different processes of speech perception, specially in the presence of multi-sounds.

Correspondencia: Adrián Fuente C.
Email: afuente@hkusua.hku.hk

INTRODUCCIÓN

En la vida diaria para poder ser capaces de seguir una conversación se requiere segregarse los sonidos del ambiente. Este fenómeno es altamente complejo debido a la presencia de muchos hablantes y diversos sonidos no verbales que simultáneamente enmascaran la voz del emisor a quien queremos atender. Sin embargo, a partir de distintas claves ya sean estas binaurales o monoaurales, el oyente es capaz de comprender lo que el hablante está diciendo. Este hecho ha sido ampliamente estudiado por muchos autores. Cherry (1953) fue el primero en abordar este fenómeno, llamándolo *cocktail party phenomenon*. Con posterioridad, diversos investigadores (Conway, Cowan y Bunting, 2001; Darwin, y Bethell-Fox, 1977; Darwin y Hukin, 1999; Freyman, Helfer, McCall, y Clifton, 1999) han tratado de identificar los principales componentes que permiten a los oyentes tanto segregarse como agrupar los sonidos¹. Así, se han estudiado aspectos como la frecuencia fundamental, diferencias interaurales de tiempo, la prosodia en el texto del hablante e incluso la memoria de trabajo.

El hecho de conocer los procesos de segregación y agrupación de los sonidos es de suma transcendencia para los clínicos del área fonoaudiológica, ya que tiene directa relación con la evaluación del procesamiento auditivo central y con las estrategias de rehabilitación para impedidos auditivos.

En la actualidad la evaluación del procesamiento auditivo central se ha transformado en un procedimiento rutinario en países anglosajones. No obstante, aún existe desconocimiento entre los clínicos acerca de los procesos auditivos subyacentes a las pruebas utilizadas.

Por otra parte, en lo que se refiere a la adaptación de audífonos, los audiólogos buscan reinsertar social y laboralmente al impedido auditivo. El audífono ayuda al sujeto a detectar mejor los sonidos y a discriminar el habla en situaciones reales como es el caso del fenómeno de *cocktail party*. Por tanto, se requiere que los audiólogos conozcan los mecanismos subyacentes del sistema auditivo en tareas de segregación de sonidos ante la presencia de ruidos enmascarantes.

En la presente revisión se abordarán aspectos y mecanismos relacionados a cómo los oyentes son capaces de segregarse los sonidos en situaciones de múltiples hablantes. Además, se discutirá cuál es el papel que juega la atención en la habilidad de mantener un sonido como el "sonido objetivo"². Finalmente, se explicarán algunas pruebas de procesamiento auditivo central que están relacionadas con la teoría de percepción del habla.

DESARROLLO

Percepción del habla bajo la presencia de sonidos enmascarantes de fondo

Habla y sonidos enmascarentes

De acuerdo a la literatura, los oyentes necesitan mantener una relación señal ruido para ser capaces de discriminar correctamente el habla. Esta relación señal ruido (S/R), a nivel del *output* de los filtros auditivos cocleares³, debería ser igual o mejor a -4 dB (Moore, 2003). Es así que

¹ Segregar sonidos son los distintos procesos que el sistema auditivo realiza para separar los distintos estímulos auditivos con el objetivo de realzar el sonido que el oyente quiere prestar atención. Agrupar sonidos son los distintos procesos que el sistema auditivo realiza para juntar las distintas características acústicas de un sonido, (características tales como intensidad y espectro frecuencial). Lo anterior al igual que la segregación de sonidos, tiene el objetivo de realzar el sonido que el oyente quiere prestar atención.

² Sonido objetivo es un término que aparecerá constantemente en la presente revisión. Dicho término se refiere al sonido que el oyente desea atender voluntariamente, segregándolo de los otros sonidos del ambiente, ya sean estos verbales o no verbales. Otros términos de igual significado son: habla objetivo, voz objetivo y oración objetivo.

³ El *output* de los filtros auditivos corresponde al material auditivo que ya ha sido analizado y filtrado frecuencialmente a lo largo de la membrana basilar—este análisis frecuencial ocurre en las primeras etapas del procesamiento auditivo—. Los componentes sinusoidales del sonido son separados y codificados independientemente en el nervio auditivo.

oyentes y hablantes (por razones de *feed-back*) normalmente ajustan su esfuerzo vocal, la distancia y orientación de la cabeza con el objetivo de mantener la relación S/R requerida. Por otra parte, cuando se desea determinar la contribución total de las voces interferentes o enmascaradoras (murmullo) en reuniones sociales, se debe considerar a todos los hablantes y a la acústica del ambiente. Plomp (1977) planteó que la relación S/R en las típicas condiciones de *cocktail-party* es 0 dB, cuando la voz objetivo y la interferente presentan el mismo nivel y el hablante objetivo está a una distancia de 0.7 metros del oyente. Habitualmente, esto es suficiente para una adecuada inteligibilidad del mensaje.

No obstante, hay situaciones en que el habla puede llegar a ser marcadamente ininteligible. Algunas de ellas son: lugares en que la acústica es deficiente y/o hay presencia de ruidos adicionales, el hablante evidencia una inadecuada fonación o pobre pronunciación o bien el oyente presenta algún impedimento auditivo (Bronkhorst, 2000).

Otros sonidos enmascarantes son los mensajes competitivos⁴ que pueden provenir de distintas fuentes y, por lo tanto, el impacto en el oyente va a diferir. Esto se traduce en diferencias físicas entre los sonidos tales como espectro frecuencial, duración e intensidad. Tales diferencias, de acuerdo al tipo de sonido enmascarante que se trate, son utilizadas por el sistema auditivo como claves para segregar y agrupar los distintos sonidos. Dependiendo de la situación de escucha, estas claves pueden ser mono o binaurales. El clásico fenómeno de *cocktail-party* considera al oyente en una condición de escucha binaural.

En cuanto al tipo de enmascarador, Miller (1947) obtuvo diferentes puntajes para el reconocimiento de la palabra dependiendo del enmascarador utilizado. Para ello, usó discurso continuo de uno o más hablantes y ruido blanco. Sus hallazgos le permitieron concluir que una voz interferente es considerablemente menos efectiva como enmascarador en comparación al ruido blanco a una misma intensidad.

Así, Miller encontró que al nivel del 50% de inteligibilidad (SRT), hubo una diferencia en casi 10 dB entre discurso y ruido blanco utilizados como enmascaradores. Esto es, el SRT obtenido con la presencia de ruido blanco fue mayor en comparación al SRT obtenido con la presencia de discurso.

Sin embargo, investigaciones posteriores han demostrado que cuando varias voces son utilizadas como enmascarador, el promedio de fluctuaciones decae y la señal se hace más bien continua en términos frecuenciales y de intensidad (Bronkhorst, 2000). Esto último torna a este tipo de enmascarador más efectivo.

Las diferencias en la eficacia de los enmascaradores entre sonidos estacionarios y de habla fueron estudiadas luego de los primeros reportes de Miller. Carhart, Tillman y Greetis (1969) encontraron una diferencia en los umbrales de inteligibilidad para palabras espondeicas⁵ de casi 4 dB utilizando habla y ruido blanco como enmascaradores.

Duquesnoy (1983) utilizó una prueba de inteligibilidad de oraciones para comparar la eficacia de enmascaramiento de una voz interferente tocada hacia atrás con un ruido de igual promedio espectral. Los resultados mostraron una diferencia de 7 dB entre los umbrales de recepción del habla (SRT) obtenidos bajo la presencia de ruido y habla enmascarante, aunque ambos enmascaradores presentaban igual espectro. También Peters, Moore y Baer (1998) parearon el espectro del ruido y del habla interferente, encontrando diferencias en los SRT de 8 dB. Esta mejoría en la

⁴ Sonidos competitivos se refiere a todos aquellos estímulos acústicos, ya sean estos verbales o no verbales, que están presentes en una situación dada y que interfieren con el sonido que el oyente desea prestar atención (sonido objetivo). Cuando se quiere hacer referencia a que estos sonidos competitivos representan mensajes verbales, se suele utilizar el término "mensaje competitivo".

⁵ Palabras espondeicas son aquellas compuestas a partir de dos palabras monosilábicas. Este tipo de palabras existen en gran número en idioma inglés. Por ello, estas palabras son ampliamente utilizadas en pruebas audiológicas relacionadas con discriminación de habla y procesamiento auditivo central. Ejemplos de estas palabras son: *cowboy*, *milkman*, *barman*.

inteligibilidad causada por las fluctuaciones del habla desaparece cuando el número de hablantes es aproximadamente cuatro o más (Bronkhorst & Plomp, 1992; Miller, 1947).

Por otra parte, la inteligibilidad decae cuando el hablante objetivo y el enmascarante son del mismo sexo (Stubbs & Summerfield, 1990).

¿Por qué sucede esta diferencia en la inteligibilidad? Los sujetos auditivamente normales son capaces de utilizar los quiebres o fluctuaciones que se producen en la voz competitiva para captar el mensaje proveniente del hablante objetivo. Estos quiebres pueden ser de dos tipos: temporales y espectrales.

Los quiebres temporales son producidos cuando la intensidad promedio del habla competitiva decae, por ejemplo, al producirse breves pausas en el discurso o durante la producción de sonidos de baja energía como el caso de los fonemas /m/, /n/, /k/ o /p/. Durante estos quiebres temporales la relación S/R es alta y ello permite por breves lapsos captar más fácilmente la señal objetivo.

Los quiebres espectrales son producidos porque el espectro del habla objetivo es usualmente diferente del habla enmascarante medido en cualquier intervalo breve de tiempo. Aun cuando partes del espectro del habla objetivo pueden ser enmascaradas por completo por el habla enmascarante, otras son difícilmente enmascaradas del todo. La relación S/R en estos puntos puede usualmente exceder 20 dB. Es así que partes del espectro del habla objetivo pueden ser fácilmente captadas y utilizadas para inferir la estructura del sonido completo de habla.

Esto último se relaciona directamente con el concepto de redundancia del lenguaje. Los oyentes no requieren percibir todos los fonemas presentes en el mensaje para comprenderlo ya que completan los vacíos (restauración fonémica) utilizando el conocimiento de la lengua, contexto e incluso el conocimiento acerca del propio hablante. Esto representa lo que algunos autores han llamado influencias *top-down* en la percepción del habla (Ryalls, 1996).

Otro factor que podría contribuir a la inteligibilidad de una señal acústica en presencia de enmascaradores de fondo es el fenómeno llamado *comodulation masking release* (CMR).

Moore (2003) mencionó que los modelos basados en CMR asumen que el sistema auditivo compara los patrones de modulación del promedio espectral a nivel del *output* de los filtros auditivos sintonizados en diferentes centros de frecuencia. Para un enmascarador comodulado sin una señal, el patrón de modulación es similar para todos los filtros auditivos que se encuentran activos. Cuando a este enmascarador se le agrega una señal, el patrón de modulación a nivel del *output* auditivo para la frecuencia sintonizada con esta señal se altera. Por lo tanto, la presencia de la señal es indicada por una disparidad en el patrón de modulación a través de los distintos filtros auditivos.

El sistema auditivo puede ser sensible a esta disparidad. Sin embargo, algunos autores como Festen (1993) han concluido que el CMR tiene solamente una influencia menor y que el hecho de escuchar a partir de diferencias temporales y espectrales entre el ruido enmascarante y el habla es el efecto dominante.

Claves monoaurales y binaurales en la discriminación del habla en presencia de sonidos enmascarantes

Para el agrupamiento auditivo y la atención selectiva se ha demostrado que los sujetos utilizan claves provenientes de la frecuencia fundamental (Darwin & Bethell-Fox, 1977) y claves espaciales como diferencias interaurales de tiempo DIT⁶ (Darwin & Hukin, 1999). Darwin y Hukin (1999) demostraron que los oyentes pueden utilizar claves interaurales de unas pocas décimas de segundo para discriminar una señal verbal en presencia de otra competitiva. Los autores encontra-

⁶ Las DIT son claves utilizadas solo en condiciones de escucha binaural, esto es, el clásico fenómeno de *cocktail party*. No está claro aún cómo es exactamente la combinación de claves mono y binaurales de las cuales el sistema auditivo utiliza en condiciones de escucha binaural.

ron que los oyentes mostraron un mínimo uso de la frecuencia fundamental para ejecutar esta tarea. Esto último equivale a que las DIT jugarían un papel más importante que la frecuencia en la atención selectiva del material auditivo verbal.

Los mismos autores en otro estudio sobre atención auditiva (Darwin y Hukin, 2000), aportaron nueva información acerca de los roles relativos de las claves espaciales, prosódicas y de características individuales de la voz (tamaño del tracto vocal).

Sus experimentos mostraron que las variaciones naturales de la prosodia y el tono vocal fueron efectivas para reconocer qué palabra correspondía a la oración escuchada previamente. No obstante, estas diferencias no serían tan importantes como las diferencias interaurales de tiempo.

En otra situación experimental de Darwin y Hukin (2000) además de variar la prosodia y el tono vocal, manipularon el tamaño del tracto vocal de los hablantes en aproximadamente 15%. Concluyeron que al existir diferencias prosódicas y de tracto vocal entre los hablantes, estas presentaron un rol más importante en la atención auditiva que las diferencias interaurales de tiempo. Sin embargo, cuando las diferencias en el tracto vocal no son significativas (menor a un 15% aproximadamente) y aun existiendo diferencias prosódicas entre los hablantes, las diferencias interaurales de tiempo serían las claves utilizadas por los oyentes para atender al sonido objetivo.

Por otra parte, Cherry (1953) sugirió que la separación espacial es el mayor contribuyente para resolver el problema del fenómeno de *cocktail-party*. Sin embargo, las claves espaciales son solo un aspecto de posibles claves utilizadas para determinar las fuentes del sonido. Por tanto, las claves espaciales no serían suficientes para la ubicación de la fuente sonora. De este modo, Yost, Dye & Sheft (1996), sugirieron que la separación espacial puede no representar la clave principal para determinar la fuente sonora. Así, después de analizar los resultados de su investigación, concluyeron que no existe una gran ventaja provista por la separación espacial cuando se trata de dos fuentes sonoras. Los autores sugirieron que las otras claves que diferencian a los hablantes (frecuencia fundamental, vibrato, prosodia) permiten a los oyentes identificar las fuentes sonoras. Lo anterior significa que entregar claves espaciales sería de poca ayuda complementaria para los oyentes.

No obstante, Yost, Dye & Sheft (1996) mencionaron que cuando el escenario acústico se vuelve más complejo con la coexistencia de tres o más sonidos, las claves espaciales adquieren un rol más importante en la identificación de las distintas fuentes sonoras. En relación a esto último, los investigadores encontraron que el rendimiento en tareas de identificación de la fuente sonora que implica el reconocimiento de estímulos auditivos, mejoró entre los sujetos de su estudio a medida que la distancia entre los emisores del habla aumentó.

Un hecho importante de señalar es que una de las condiciones estudiadas, se les permitió a los oyentes mover la cabeza para localizar e identificar de mejor forma la fuente sonora. Los sujetos mencionaron que estos movimientos los ayudaron a identificar las fuentes sonoras. Sin embargo, los autores sugirieron que los movimientos de cabeza jugarían un rol más importante en tareas de localización que de identificación.

Resultados similares fueron encontrados por Hawley, Itsovky y Colburn (1999) en un estudio de inteligibilidad y localización del habla en un ambiente de múltiples fuentes sonoras. Los autores encontraron que la proximidad de las oraciones competitivas a la ubicación de las oraciones objetivo influyó más que el número de fuentes en la inteligibilidad de las oraciones bajo condiciones binaurales y de escucha por el mejor oído. Bajo condiciones monoaurales y de escucha por el peor oído, el número de sonidos competitivos influyó más en el rendimiento de inteligibilidad que la proximidad entre las oraciones competitivas y las objetivo. Por lo tanto, la configuración espacial parece tener un efecto más importante que el hecho de agregar o remover una fuente sonora.

En cuanto al rol de los enmascaradores se describen dos tipos de sonidos enmascaradores: *energéticos* y *de información*.

Un enmascarador energético es aquel que se sobrepone al sonido objetivo en tiempo y espectro frecuencial. Es decir, un sonido que es capaz de enmascarar a través de sus características físicas.

Un enmascarador de información se produce cuando tanto el sonido objetivo como el enmascarador son audibles por el oyente. Es decir, existe una relación S/R suficiente para escuchar ambos sonidos. El efecto del enmascarador de información recae a través del contenido lingüístico de este, y por tanto, lo que el oyente necesita es mantener activamente su atención en el hablante objetivo, eliminando el o los hablantes enmascaradores. Ciertamente, muchos enmascaradores actúan tanto como sonidos enmascaradores energéticos y de información.

Atención y memoria en tareas de percepción del habla

Existen dos procesos principales involucrados en la percepción del habla bajo la presencia de sonidos enmascarantes de fondo: *atención dividida* y *atención selectiva*. La atención dividida se produce cuando el sujeto requiere obtener información proveniente de dos o más fuentes sonoras. Por el contrario, la atención selectiva ocurre cuando el oyente necesita concentrarse en solo una fuente sonora, segregándola de las otras.

Algunos estudios han reportado que durante tareas de atención selectiva, los sujetos realizan un cambio de atención hacia un canal hasta ese momento desatendido (Cherry, 1953, Wood & Cowan, 1995). En la mayoría de los casos cuando estas modificaciones ocurren, los sujetos fallan en la ejecución de la tarea principal unos pocos milisegundos después de producido el cambio en el foco de atención. Se plantea que las tareas de atención selectiva requerirían de la memoria de trabajo ya que la función de esta es mantener activamente la información relevante al servicio de tareas cognitivas complejas (Miyajima & Shah, 1999). Además, se ha relacionado la generación de estos cambios en la atención con un bajo rendimiento en la memoria de trabajo (Conway, Cowan & Bunting, 2001).

En un estudio, Conway *et al.* (2001) encontraron que el 65% de los sujetos con un bajo *span*⁷ fallaron en desatender el mensaje irrelevante que no tenía que ser atendido, en comparación al 20% de los sujetos con un *span* elevado. Además, los sujetos con bajo *span* reportaron mayores dificultades en disminuir la atención sobre el mensaje irrelevante.

Los autores sugirieron que las tareas de *span* son esencialmente tareas de atención dividida, mientras que otras tareas como las de escucha dicótica son de atención selectiva.

Parece ser que existe una habilidad cognitiva general que permite a los sujetos rendir correctamente en diferentes tareas. Conway *et al.* (2001) consideraron que esta habilidad está íntimamente ligada a la capacidad de la memoria de trabajo.

Por otra parte, de acuerdo a la teoría de atenuación de Treisman (1964), el material del canal acústico desatendido hace contacto con representaciones de la memoria de largo plazo y activaría estas representaciones más bien a un nivel consciente. Cambios en el patrón del canal de los estímulos irrelevantes para el oyente, son captados si originan un desajuste en el nivel de atención de ese canal. El hecho que el sujeto reacomode su nivel de atención posiblemente depende de la magnitud de la modificación del estímulo en sus características físicas y en el nivel preexistente de activación de las unidades en la memoria (estimuladas después de producido un ajuste en el nivel de atención).

El procesamiento automático que produce tal reacomodación en la atención, puede ser explicado por la teoría de respuesta orientada de Sokolov⁸ (1963). De acuerdo a esta teoría, los

⁷ *Span* se traduce literalmente en español como extensión. El término *span* es ampliamente utilizado en psicología experimental relacionada con la exploración de tareas de memoria de trabajo. *Span* se refiere a un número total de estímulos que el sujeto es capaz de retener luego de realizada una tarea. Por ejemplo, se presentan en una pantalla distintos estímulos visuales correspondientes a diferentes campos semánticos y luego de terminada la presentación, se le solicita al sujeto verbalizar todas los estímulos que recuerde.

⁸ La respuesta orientada fue originariamente descrita por Iván Pavlov, quien la denominó como "reacción investigativa" (Nätäänen, 1992). Pavlov la describió: Es el reflejo que lleva consigo la respuesta inmediata en los hombres y animales a los más leves cambios en el mundo que los rodea, de tal forma que ellos puedan inmediatamente orientar el órgano receptor apropiado de acuerdo a la cualidad perceptiva del agente que lleva consigo el cambio, de tal forma de hacer una completa investigación de este (Pavlov, 1927, P:12).

sujetos forman una representación en la memoria o un “modelo neural” de cada estímulo de su medio ambiente. Para que las características de un nuevo estímulo puedan ser analizadas, sus representaciones se comparan con las representaciones preexistentes en la memoria. Una diferencia entre los dos produce una respuesta orientada.

Aun cuando el análisis de un estímulo irrelevante probablemente no puede producir un completo set de características para compararlo con el modelo, parece claro, al menos, que algunos análisis de las características físicas ocurren automáticamente (Cowan, 1988; Näätänen, 1992). Por tanto, un cambio desde habla acelerada a habla enlentecida podría resultar automáticamente en una discrepancia entre las representaciones existentes antes y después de la modificación en los estímulos de habla del canal irrelevante, elicitando de tal forma un cambio de atención en ese canal.

Al procesamiento que ocurre sin intención, incluso contra la intención, los autores lo refieren como el efecto Stroop (Näätänen, 1992). Este efecto provee sólida evidencia que el sujeto no siempre puede evitar el procesamiento de aspectos de un *input* que desea ignorar. Un ejemplo de ello es la activación semántica que ocurre sin estar conscientes. Posner y Snyder (1975) señalaron que los resultados de un número de estudios de escucha dicótica concluyen que una gran cantidad de procesamiento semántico automático ocurre en el canal que eventualmente es ignorado a nivel consciente.

En resumen, tanto la atención como la memoria de trabajo son necesarios para el reconocimiento del sonido objetivo bajo la presencia de sonidos enmascarantes. Estos procesos cognitivos son especialmente importantes para segregar sonidos y focalizarse en el que el oyente desea atender bajo la presencia de enmascaradores de información.

Pruebas de procesamiento auditivo central y los procesos auditivos involucrados

En esta sección, brevemente se discutirán algunas pruebas que evalúan el procesamiento auditivo central y su correlación comportamental. Esto es, los procesos que subyacen a la tarea solicitada. Se abordarán las pruebas de escucha dicótica, procesamiento temporal, tareas monoaurales de baja redundancia e interacción binaural. De acuerdo a la American Speech-Language-Hearing Association (1996), estos son los aspectos que deben ser considerados cuando se evalúa el procesamiento auditivo central.

Escucha dicótica

En las pruebas de escucha dicótica pueden ser utilizados distintos tipos de estímulos (dígitos, palabras espondeicas, oraciones, sonidos ambientales, sílabas, entre otros). Clásicamente se han descrito dos tareas en las pruebas de habla dicótica (Chermark y Musiek, 1997): Separación binaural e integración binaural.

En la separación binaural se le solicita al sujeto repetir solo los estímulos de un oído, esto es, atención selectiva. Por lo tanto, la memoria de trabajo es uno de los principales procesos involucrados. En la integración binaural el sujeto debe repetir los estímulos presentados en ambos oídos lo que implica la atención dividida.

Pruebas dicóticas clásicas para atención selectiva son las oraciones competitivas y la identificación de oraciones sintéticas con mensaje contralateral competitivo (Chermark & Musiek, 1997).

También puede ser utilizada la prueba de dicótico de dígitos. En estas pruebas dos estímulos diferentes son presentados para cada oído al mismo tiempo. De esta forma existe un efecto de separación espacial entre los estímulos. En estas pruebas el sonido enmascarante actúa más bien como enmascarador de información. Por lo tanto, la separación espacial percibida sirve como clave para el oyente con la finalidad de segregar los estímulos auditivos (Freyman, Helfer, McCall & Clifton, 1999; Kidd, Mason, Rohtla & Deliwala, 1998).

Estas pruebas pueden ser usadas ya sea con fonos o parlantes. Cuando se utilizan parlantes, se debe contar con una sala anecoica, ya que en salas convencionales las reflexiones pueden afectar las diferencias interaurales que forman la base de la ventaja de la separación espacial (Freyman, Balankrishnan & Helfer, 2001).

Pruebas para atención dividida son palabras espondeicas alternadas (*staggered spondaic words*) y la prueba de dicótico de dígitos, entre otras. La prueba de palabras espondeicas alternadas considera la presentación de cuatro palabras por ítem, dos palabras para cada oído. De estas cuatro palabras, dos se presentan en condición dicótica y dos en condición no dicótica. La tarea que el sujeto debe realizar es repetir las cuatro palabras en el mismo orden de presentación. El proceso aditivo involucrado en esta tarea es el de integración biaurol.

Procesos temporales

Se ha destacado que el procesamiento temporal es un componente clave de la función auditiva central (Pinheiro y Musiek, 1985). Existen distintas pruebas para evaluar diferentes aspectos del procesamiento temporal. De acuerdo a Emmanuel (2002), una de las pruebas más ampliamente utilizadas entre los clínicos en Estados Unidos es la secuencia de patrones de altura *-Pitch Pattern Sequence-*. El sujeto en esta prueba debe señalar una secuencia de tres estímulos de acuerdo a su altura y al orden de presentación.

Para realizar esta prueba es necesario usar una cámara silente para evitar señales competitivas. El sujeto debe ser capaz de detectar, discriminar y reconocer primeramente cada uno de los componentes y luego la secuencia completa (los tres estímulos). Sin duda una variable a considerar en el rendimiento de esta prueba es la memoria de trabajo.

Una prueba similar es la secuencia de patrones de duración *-Duration Pattern Sequence-*. Esta prueba utiliza diferencias de duración entre los estímulos.

En términos generales ambas pruebas evalúan habilidades similares y en ellas se destaca el traspaso de la información a través del cuerpo caloso. Eventualmente la información debe alcanzar el hemisferio derecho para su reconocimiento. Luego, debe ser procesada en este hemisferio y pasar al hemisferio izquierdo para su nominación. Masquelier (2003) señala que esta prueba evalúa la coordinación de información suprasegmental con la información segmental a través del cuerpo caloso.

Tareas monoaurales de baja redundancia

En este grupo se encuentran aquellas pruebas que contemplan habla bajo la presencia ya sea de ruido enmascarante (habla en el ruido) o mensajes competitivos (identificación de oraciones sintéticas). Además en esta categoría están las pruebas que utilizan una señal de habla degradada (habla filtrada, habla comprimida). Estas pruebas se presentan monoauralmente.

Para todas aquellas pruebas que utilizan una señal competitiva, los procesos involucrados son aquellos discutidos en la primera parte del artículo, los cuales se refieren a la segregación y agrupación de los estímulos auditivos. A continuación se expondrán con más detalle algunas de las pruebas que involucran tareas monoaurales de baja redundancia.

Habla en el ruido: Esta prueba usualmente utiliza ruido blanco como sonido interferente, esto es, un enmascarador energético. Para establecer figura y fondo auditivo, es decir, segregación de los sonidos, el oyente requiere usar claves monaurales. El espectro entre el ruido y el habla ciertamente no es el mismo, por tanto, una de las claves a utilizar por el sistema auditivo es la diferencia en el espectro frecuencial. La membrana basilar provee las bases iniciales para el proceso de filtración requerido. Es así como luego a nivel del *output* de los distintos filtros auditivos se realizan comparaciones del material procesado con la finalidad de resaltar la detección de la señal (Moore, 2003), en este caso las palabras.

Identificación de oraciones sintéticas: Esta prueba utiliza habla como señal objetivo y como señal competitiva. El oyente debe hacer uso de diferentes claves para lograr discriminar las oraciones. Diferencias en la frecuencia fundamental de los hablantes es una de las claves a utilizar. Además, la prosodia, acento e incluso intensidad pueden ser usados como claves por el sistema auditivo.

Algunos estudios han encontrado que los oyentes utilizan las diferencias en los niveles de intensidad de dos hablantes para atender selectivamente al hablante cuya voz es de menor intensidad (Brungart, 2001).

Además, el oyente puede hacer uso de las fluctuaciones de frecuencia e intensidad en el habla enmascarante. Por tanto, como ya fue discutido, el sistema auditivo puede utilizar los quiebres producidos por las fluctuaciones del habla enmascarante para obtener información del habla objetivo.

Cabe destacar que el habla enmascarante en esta prueba actúa tanto como enmascarador energético (el cual dependerá de la relación S/R) y enmascarador de información. Por lo tanto, la atención y la memoria de trabajo son dos procesos de suma transcendencia para la interpretación de los resultados.

Para ambas pruebas, habla en el ruido e identificación de oraciones sintéticas, los sujetos pierden información del hablante objetivo. Es decir, requieren cerrar la información, lo que se denomina cierre auditivo. Este proceso se lleva a cabo a través de la redundancia de la lengua, esto es, las influencias *top-down* involucradas en la percepción del habla (Ryalls, 1996).

Interacción biaural

Básicamente, la interacción biaural se refiere a aquellas tareas donde la información alcanzada por ambos oídos debe interactuar. Diferencias en el nivel de enmascaramiento para tonos y habla son tareas de interacción biaural que pueden ser utilizadas en la evaluación del procesamiento auditivo central (Schoeny & Talbott, 1994). No obstante, los clínicos no evalúan frecuentemente con este tipo de tareas (Emmanuel, 2002).

La prueba de la diferencia en el nivel de enmascaramiento (*Masking level difference* en inglés con sus siglas MLD) mide la diferencia en el umbral biaural para tonos o habla en el ruido (enmascarador) que están en fase o fuera de fase entre los oídos. Específicamente, cuando la señal está fuera de fase (antifásica) entre los oídos, existe una liberación del enmascaramiento y los umbrales de detección de la señal mejoran en comparación a la condición fásica (Chermark & Musiek, 1997).

Masquellier (2003) señala que la interacción biaural es lo que permite la localización y lateralización del sonido. Además, el autor plantea que la interacción biaural se relaciona con la detección de señales en el ruido y con la fusión biaural. Este último término es el proceso auditivo que explica que la información proveniente de ambos oídos se junta, a nivel de las estructuras del puente en el tallo cerebral, para conformar un solo estímulo. La prueba que mide este proceso es denominada "fusión biaural".

DISCUSIÓN

Distintos modelos han sido desarrollados con el objetivo de explicar cómo los oyentes son capaces de utilizar claves monoaurales y biaurales para segregar y agrupar los sonidos. Aún está en discusión cuáles son las claves (frecuencia fundamental, diferencias interaurales de tiempo, prosodia) más importantes para realizar estas tareas.

Además, procesos cognitivos como la atención y memoria de trabajo juegan un rol crucial en tareas de escucha dicótica y en la discriminación del material auditivo verbal bajo la presencia de enmascaradores de información.

Ciertamente, es fundamental que los clínicos al momento de evaluar el procesamiento auditivo central, estén conscientes de aquellas variables que pueden influir en el rendimiento del paciente como la memoria de trabajo.

Por otra parte, se han planteado los principales procesos involucrados en el rendimiento de algunas pruebas de procesamiento auditivo central. El fin de lo anterior es relacionar la percepción de habla en situaciones desfavorables con la evaluación del procesamiento auditivo central.

El debate sobre los fundamentos de la percepción de habla y su directa relación con el procesamiento auditivo central aún no está concluido. Sin duda, se requiere mayor investigación en esta área para comprender mejor cómo el sistema es capaz de segregar y agrupar los sonidos en situaciones de varios hablantes.

Finalmente, a pesar del gran desarrollo del procesamiento auditivo central en países anglosajones, en países hispanoparlantes ha sido un área de escaso desarrollo. Muchas de las pruebas que han sido descritas en el presente artículo, están disponibles solo para poblaciones angloparlantes, ya sea porque utilizan material verbal, o bien, porque las normas han sido obtenidas para estas poblaciones.

Lo anterior representa un desafío para los psicólogos experimentales y para los clínicos en el área de la audiología y de los trastornos del lenguaje. Indiscutiblemente, contar con pruebas de procesamiento auditivo central para poblaciones hispanoparlantes permitiría a los clínicos que trabajan en estos países identificar aquellos procesos auditivos que pueden dar cuenta o relacionarse con trastornos de la comunicación.

BIBLIOGRAFÍA

- AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION. (1996). Central auditory processing: Current status of research and implications for clinical practice. *American Journal of Audiology*, 5(2), 41-54.
- BRONKHORST, A. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica*, 86, 117-128.
- BRONKHORST, A., & PLOMP, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the acoustical society of America*, 92, 3132-3138.
- BRUNGART, D. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the acoustical society of America*, 109, 1101-1109.
- CARHART, R., & TILLMAN, T.W., GREETIS, E.S. (1969). Perceptual masking in multiple sound background. *Journal of the Acoustical Society of America*, 45, 694-703.
- CHERMAK, G.D., & MUSIEK, F.E. (1997). Central auditory processing disorders. New perspectives. San Diego: Singular.
- CHERRY, E.C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25, 975-979.
- CONWAY, A.R.A., COWAN, N., & BUNTING, M.F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8, 331-335.
- COWAN, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin*, 104, 163-191.
- DARWIN, C.J., & BETHELL-FOX, C.E. (1977). Pitch continuity and speech source attribution. *Journal of Experimental Psychology. Human Perception and Performance*, 3, 665-672.
- DARWIN, C.J., HUKIN, R.W. (1999). Auditory objects of attention: The role of interaural time-differences. *Journal of Experimental Psychology. Human Perception and Performance*, 25, 617-629.
- DARWIN, C.J., & HUKIN, R.W. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *Journal of the Acoustical Society of America*, 107, 970-977.
- DUQUESNOY, A.J. (1983). Effect of a single interfering noise or speech source on speech intelligibility. *Journal of the Acoustical Society of America*, 74, 739-743.
- EMMANUEL, C.D. (2002). The auditory processing battery: Survey of common practices. *Journal of the American Academy of Audiology*, 13, 93-117.
- FESTEN, J.M. (1993). Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice. *Journal of the Acoustical Society of America*, 94, 1295-1300.
- FREYMAN, R.L., Balankrishnan, U., & Helfer, K.S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America*, 109, 2112-2122.

- FREYMAN, R.L., HELFER, K.S., MCCALL, D.D., & CLIFTON, R.K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America*, 106, 3578-3588.
- HAWLEY, M.L., ITOVSKY, R.Y., & COLBURN, H.S. (1999). Speech intelligibility and localization in a multi-source environment. *Journal of the Acoustical Society of America*, 105, 3436-3448.
- KIDD, J.G., MASON, C.R., ROHTLA, T.L., & DELIWALA, P.S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America*, 104, 422-431.
- MASQUELIER, M-P. (2003). Management of auditory processing disorders. *Acta Oto-Rhino-Laryngologica Belg*, 57, 301-310.
- MILLER, G.A. (1947). The masking of speech. *Psychological Bulletin*, 44, 105-129.
- MIYAJE, A., & SHAH, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.
- MOORE, B.C.J. (2003). *An introduction to the psychology of hearing (Fifth Ed)*. San Diego: Academic Press.
- NÄTÄÄNEN, R. (1992). *Attention and brain function*. New Jersey: Lawrence Erlbaum Associates.
- PAVLOV, I.P. (1927). *Conditioned Reflexes (G.V. Anrep, Trans)*. London: Pergamon Press.
- PETERS, R.W., MOORE, B.J.J., & BAER, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Journal of the Acoustical Society of America*, 103, 577-587.
- PINHEIRO, M.L., & MUSIEK, F.E. (1985). *Assessment of central auditory dysfunction: Foundations and clinical correlates*. Baltimore: Williams & Wilkins.
- PLOMP, A. (1977). Acoustical aspects of cocktail parties. *Acustica*, 38, 186-191.
- POSNER, M. I., & SNYDER, C. R. R. (1975). Attention and cognitive control. En R. L. Solso (Ed), *Information processing and cognition, The Loyola Symposium*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- RYALLS, J. (1996). *A basic introduction to speech perception*. Singular: San Diego.
- SCHOENY, Z., & TALBOTT, R (1994). Non-speech procedures. En central testing. In J. Katz (Ed), *Handbook of Clinical Audiology (Fourth ed., pp. 212-221)*. Baltimore: Williams & Wilkins.
- SOKOLOV, Y.N. (1963). *Perception and the conditioned reflex (W. Waydenfeld, Trans)*. London: Pergamon Press.
- STUBBS, R.J., & SUMMERFIELD, Q. (1990). Algorithm for separating the speech of interfering talkers: Evaluation with voiced sentences and normal-hearing and hearing impaired listeners. *Journal of the Acoustical Society of America*, 87, 359-372.
- TREISMAN, A.M. (1964). Selective attention in man. *British Medical Bulletin*, 20, 12-16.
- WOOD, N.L., & COWAN, R.W. (1995). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology. Learning, Memory, & Cognition*, 21, 255-260.
- YOST, W.A, DYE R.H., & SHEFT, S. (1996). A simulated "cocktail party" with up to three sound sources. *Perception & Psychophysics*, 58, 1026-1036.