

CAPÍTULO 11

LOS EFECTOS DEL RUIDO SOBRE LA PERCEPCIÓN DEL HABLA. APLICACIONES AUDIOMÉTRICAS

VICTORIA MARRERO
UNED

MARCELO J. RODRÍGUEZ CRUZ
ALFONSO IGUALADA PÉREZ
CSIC

1. INTRODUCCIÓN

1.1. *Los ruidos y sus efectos*

1.1.1. El ruido en el sistema de la comunicación

Desde un punto de vista físico, no hay ninguna diferencia entre el ruido y cualquier otro sonido: en todos los casos se trata de movimientos en las partículas del aire que nos rodea; lo que convierte un sonido en ruido es un rasgo psicológico: “el ruido se define como un sonido indeseado” (Namba, Kuwano y Schick, 1986). Por lo tanto, en todo estudio sobre el ruido resulta inevitable considerar la percepción subjetiva de ese estímulo por parte del receptor; si nuestra atención se centra, además, en la percepción del habla en entornos ruidosos, podemos comenzar por situar el ruido dentro del sistema de la comunicación, considerándolo como la interferencia (de todo tipo, no solo sonora) que sufre la señal; el sistema utiliza los canales de información oral-auditivo y gestual-visual de forma sincronizada, generando mecanismos flexibles de compensación cuando la tarea se vuelve exigente, para asegurar el éxito de una comunicación acorde al contexto.

Limitándonos al ruido como fenómeno sonoro, su presencia en el entorno urbano es continua y creciente, hasta el punto de ser considerado un importante contaminante ambiental, con efectos adversos sobre la salud de la población, el primero de los cuales es el daño causado al sistema auditivo, especialmente en jóvenes muy expuestos a música a elevadas intensidades¹.

1.1.2. Tipos de ruido

Existen muchas taxonomías para clasificar los ruidos: algunas atienden al tipo de componentes acústicos que predominan en ellos; por ejemplo, ruido de alta o baja frecuencia (Berglund, Hassmen y Job, 1996); ruido blanco, de espectro plano; ruido rosa, de intensidad inversamente proporcional a la frecuencia; ruido gris, equisonor, etc. Otras establecen las categorías en función de su origen: ruidos comunitarios en áreas urbanas, suburbanas y rurales, en áreas de trabajo o recreativas, en aulas escolares; ruido de aviones; ruido del tráfico rodado y otros transportes de superficie, ruidos industriales, etc. (National Academy of Engineering, 2010). En el caso del habla, el tipo de ruido que mayores interferencias produce para el reconocimiento de la señal es precisamente el generado mediante habla, el que encontramos cuando diferentes conversaciones se superponen.

En este ámbito, la comunidad científica ha utilizado varios tipos de ruidos representativos del murmullo de fondo de varias personas. El ruido-multihablante, (*multitalker babble*), se genera a partir de grabaciones humanas y pretende emular al máximo las condiciones de una situación natural de ruido de varios hablantes. El oyente, con el fin de acceder a la información, debe segregar el estímulo presentado (sílabas, palabras, frases) de esas conversaciones de fondo. El nivel de inteligibilidad de ese ruido conversacional de fondo es variable, y depende, entre otros factores, del número de locutores que se haya superpuesto para crear el estímulo: por eso contamos con ruido multihablante de dos locutores, de cuatro,

¹ En Mora, Martínez y Betancourt (2007) se presentan datos de contaminación acústica en Mérida (Venezuela) y los resultados de una encuesta entre jóvenes sobre su conciencia al respecto.

de seis, del mismo sexo, de ambos... (Wilson, 2003). También se ha debatido si el ruido multihablante creado con una lengua produce los mismos efectos en otra diferente (cf. 1.2.4 y 3).

En otros estudios actuales, se opta por utilizar ruido-con-forma-de-habla (*speech-shaped noise*), que se consigue por medio de programas informáticos, mediante lo que se conoce como *separación de la fuente y el filtro*, es decir, segregando la señal originada en los pliegues vocales de la señal conformada por todas las estructuras fijas o móviles implicadas en la articulación del habla. Algunos estudios han comparado el efecto del ruido con forma de habla y del ruido multihablante sobre la señal lingüística, concluyendo que resulta más enmascarante² el segundo que el primero (Hoen *et al.*, 2007; Parikh y Loizou, 2005; Hall, Grose, Buss y Dev, 2002, llegan a conclusiones similares, especialmente en niños).

En trabajos de orientación psicolingüística, especialmente, aunque también en algunos del ámbito audiológico (Tillman, Carhart y Olsen, 1970) se utiliza *competing speech*, habla en competencia, para enmascarar la señal lingüística: en este caso, se trata de una grabación paralela, a menudo perfectamente inteligible (aunque en algunos experimentos se invierte la señal, para quitarle el contenido semántico, como en el trabajo clásico de Dirks y Bower, 1969), pero a la que el sujeto no debe prestar atención. El habla en competencia se utiliza especialmente para valorar el enmascaramiento informativo, y diferenciarlo del energético (Schneider, Li y Daneman, 2007), como veremos a continuación.

1.1.3. Efectos del ruido

El ruido provoca un efecto sobre la señal que se conoce como *enmascaramiento*, “el proceso por el cual el umbral de audibilidad de un sonido aumenta debido a la presencia de otro sonido (máscara)” (American Standards Association, 1960). Por tanto, con este término nos referimos “a la presencia de un estímulo que interfiere en la percepción de otro” (Marrero, 2008: 221).

Podemos clasificar el enmascaramiento utilizando diferentes criterios:

² Cf. más abajo, 1.1.3, el concepto de enmascaramiento.

- a) Según el orden de presentación de los estímulos, se habla de enmascaramiento simultáneo (los dos a la vez, la energía de la *máscara* se superpone a la de la señal); progresivo o regresivo (la máscara interfiere con el procesamiento de la señal: intercepta la memoria, crea un referente incorrecto, le añade información...).
- b) En cuanto a las vías de presentación, los estímulos pueden ser escuchados binauralmente, monoauralmente, o dicóticamente (un sonido diferente por cada oído).
- c) Según el tipo de estímulo utilizado: la máscara y el modelo pueden ser la misma clase de sonido (tono sobre tono) o diferentes (por ejemplo, ruido sobre tono).
- d) También puede variar la tarea a la que nos enfrentemos: identificar el estímulo o solo señalar cuándo se ha presentado, detectarlo.

Pero cuando hablamos de interferencias en la descodificación del habla por estímulos también de habla, seguramente la dicotomía más importante es la que separa el efecto físico del cognitivo: por una parte sufrimos el llamado “enmascaramiento energético”, que se produce como un fenómeno puramente auditivo, debido a que los dos estímulos presentan energía en las mismas bandas de frecuencia (este es un fenómeno bien conocido y descrito desde el punto de vista psicoacústico). Pero también puede darse el “enmascaramiento informativo”, donde elementos cognitivos de alto nivel interfieren con la capacidad del oyente para extraer un patrón de la señal enmascarada (Brungart, 2001; Freyman, Balakrishnan y Helfer, 2001). Algunos de esos elementos son el grado de familiaridad con el contenido del mensaje, la mayor o menor semejanza entre la voz de la señal y las de la máscara (ruidos producidos por locutores del mismo sexo que el del hablante producen más efecto que si son diferentes, por ejemplo); la localización del ruido y la de la señal en el escenario auditivo; la riqueza de claves visuales del habla en el locutor; y también las diferencias cognitivas entre sujetos (cf. 1.2.2.); todo ello hace que los resultados del enmascaramiento informativo sean mucho más variables y difíciles de cuantificar que los del enmascaramiento energético.

1.2. *La percepción del habla en el ruido*

1.2.1. Bases anatómicas, fisiológicas y neurológicas de la percepción del habla en entornos ruidosos

El sistema auditivo es un complejo conjunto de órganos cuya función es transformar las ondas acústicas que nos rodean en estímulos eléctricos en la corteza cerebral. En ese camino entre la oreja (o pabellón auditivo) y el cerebro la mayor parte de las vías son ascendentes (vías *aférentes*); sin embargo, contamos también con una ruta descendente, que se inicia en la corteza y llega hasta la cóclea, en el oído interno (vías *eferentes*), cuya función sigue siendo, en buena medida, un misterio (para una revisión en español, cf. Déllano, Robles y Robles, 2005). Pues bien, contamos con un conjunto de evidencias cada vez mayor que nos llevan a pensar que este “camino de retorno” desempeña un papel esencial en nuestra capacidad para identificar la señal (el habla, en nuestro caso) en entornos ruidosos; y más concretamente uno de sus núcleos, el haz o fascículo olivo-coclear medial (en inglés, *auditory medial olivocochlear; MOC*), que actuaría como un anti-enmascarador gracias a la capacidad de inhibición que le permiten sus conexiones con las células ciliadas externas (Kawase *et al.*, 1993a y 1993b).

Los experimentos que han permitido alcanzar esta conclusión se han realizado tanto con sujetos normooyentes como con personas afectadas por deficiencias, en niños, jóvenes y en mayores. Así, Giraud *et al.* (1997) midieron la inteligibilidad del habla en ruido con pacientes cuyas vías eferentes habían sido cortadas por una intervención quirúrgica y en sujetos normales, y obtuvieron un alto grado de correlación entre la efectividad del funcionamiento olivo-coclear y las tasas de inteligibilidad. En cuanto a los diferentes grupos de edad, Kumar y Vanaja (2004) estudiaron las respuestas automáticas (otoemisiones acústicas³) de diez niños con buenos resultados académicos ante la presentación de señales de habla en

³ Las otoemisiones acústicas son sonidos de baja intensidad que genera el propio oído normal como resultado de la mecánica coclear; se sabe que la estimulación acústica de una cóclea puede modificar la activación de las fibras aferentes en la cóclea del oído contrario; por tanto, es posible estudiar de forma no invasiva la actividad del haz olivo-coclear medial midiendo las otoemisiones obtenidas mientras se presenta ruido en el oído contrario (Kim, Frisina y Frisina, 2006).

silencio y con ruido a diferentes intensidades, tanto con estimulación del oído contrario como sin ella; sus resultados también confirman el papel del haz olivo-coclear medial; en el sentido inverso, Muchnik *et al.* (2004) compararon las respuestas de 15 niños con trastornos del procesamiento auditivo, en comparación con un grupo de control: entre los primeros se detectó una baja actividad del sistema olivo-coclear medial, “que indicaría una función inhibitoria reducida y afectaría a su capacidad para oír en presencia de ruido ambiente” (Muchnik *et al.*, 2004: 107). Y lo mismo encontramos en las conclusiones de Kim, Frisina y Frisina (2006) con tres grupos de adultos (jóvenes, de mediana edad y ancianos): las dificultades que las personas mayores experimentan para entender el habla en ruido se relacionan con el deterioro que provocan los años en el sistema eferente, y más concretamente del haz olivo-coclear medial, que funcionaría como “un filtro adaptativo no lineal durante el procesamiento del habla en entornos ruidosos, y también como un procesador de *cocktail party*” (cf. más abajo, 1.2.4; Kim, Frisina y Frisina, 2006: 860).

1.2.2. Claves acústicas en la identificación del habla en ruido

El ruido de habla afecta de modo diferente a unas claves acústicas u otras en la señal lingüística. Estudios sobre el inglés (Parikh y Loizou, 2005) muestran que, en el caso de las vocales, el segundo formante resultó más frágil que el primero; en cuanto a consonantes, las oclusivas presentaron unas tasas de identificación altas incluso en condiciones extremas, donde la información sobre la barra de explosión no se preservaba; los autores atribuyen a las transiciones formánticas el papel de clave principal. Si consideramos los rasgos fónicos, en un estudio con niños disléxicos a los que se presentaron pseudopalabras enmascaradas con ruido-conforma-de-habla, Ziegler *et al.* (2009) encontraron que los lugares de articulación resultaban más frágiles que los modos de articulación o las diferencias entre sonidos sonoros y sordos.

Algunos de estos resultados son coherentes con los obtenidos para el español utilizando palabras en condiciones que imitaban los efectos de una hipoacusia neurosensorial sobre la señal (filtro en frecuencias altas y reverberación; Marrero, 1990): los lugares

de articulación resultaron menos resistentes a la distorsión que los modos, pero en nuestro experimento las vocales obtuvieron mejores índices de reconocimiento que las consonantes, incluidas las oclusivas.

De esta información podemos deducir que, posiblemente debido a la redundancia, diferentes claves acústicas de los sonidos del habla, pese al efecto de ruido, hemos adquirido una representación suficientemente robusta de estos, como para poder identificarlos en situaciones de exigencia moderada. Sin embargo, cuando la exigencia es alta, se han recogido resultados diferentes. Así, en el estudio de Cunningham *et al.* (2001), con niños con patología del aprendizaje, se demostró que la modificación de las claves de las oclusivas, en concreto el incremento de la duración de espacio blanco en la fase de cierre, y el incremento de la intensidad de la barra de explosión en la fase de apertura, mejora la percepción de los sonidos del habla, al igual que las representaciones neurofisiológicas de las características degradadas por el ruido. Además, el incremento de la clave de intensidad de la barra de explosión reduce el enmascaramiento producido en la vocal siguiente.

Pasando del nivel segmental al suprasegmental, las variaciones en la frecuencia fundamental son uno de los elementos perceptivamente más potentes en el habla: intervienen de forma determinante en las variaciones prosódicas, como hemos mencionado, pero también son un elemento básico en la identificación del locutor (Rose, 2002). En el reconocimiento de habla enmascarada por habla, la frecuencia fundamental de los locutores constituye una variable importante: ya nos hemos referido a que las voces del mismo sexo se enmascaran más entre sí que las de sexos diferentes. En general, cuanto más distantes estén las frecuencias fundamentales de los hablantes, más fácil será la discriminación del mensaje. También facilita la descodificación la familiaridad con la voz de la señal (Brungart, Simpson, Ericson y Scott, 2001).

1.2.3. Claves visuales en la identificación de habla en ruido

Los estudios sobre percepción multimodal del habla han despertado un creciente interés por las relaciones entre la comunicación gestual y la verbal: se considera que el gesto es un compo-

nente integral del lenguaje que refuerza la función comunicativa y añade información al contenido verbal. Gestos y habla se sincronizan a nivel semántico, pragmático y fonológico, y en caso de claves en conflicto, la información visual puede resultar más determinante que la auditiva (Prieto *et al.*, 2011). En situaciones de comunicación con ruido el gesto no se ve afectado, por lo que su papel cobraría especial preponderancia, hasta el punto de alcanzar una “efectividad inversa” respecto al sonido: “el efecto de verle el movimiento de los labios al hablante incrementa cuando se reduce la relación señal-ruido” (Barutchu *et al.*, 2010: 39).

En conclusión, en los actos comunicativos en los que nos encontramos en el mismo espacio del interlocutor, la información de claves visuales parece tomar prominencia en condiciones de ruido.

1.2.4. Factores cognitivos

La *atención y la memoria* son componentes esenciales para el procesamiento de la información verbal. En entornos ruidosos, para extraer la información es necesario focalizar la atención en una fuente de información e inhibirla simultáneamente de los flujos informativos de fondo. La capacidad para cambiar el orden de procesamiento en los flujos informativos a los que se quiere atender se relaciona con el uso de la memoria de trabajo, en niveles centrales de procesamiento, y presenta grandes variaciones entre individuos (Schneider, Li y Daneman, 2007; los efectos distractores del ruido son más intensos para los niños pequeños que para los mayores o los adultos: Fallon, 2001).

Una de las líneas de investigación más fructíferas del último medio siglo, en este ámbito, es la relacionada con el denominado efecto *cocktail party*. Con esta denominación bautizó Collin Cherry, en 1953, la capacidad humana para focalizar la atención en un solo hablante en medio de ruido de múltiples conversaciones. Muchos de esos trabajos se han dirigido a estudiar la separación espacial de las fuentes de información, y la escucha binaural; pero otros se han centrado en las características de la señal de habla que hacen posible este fenómeno: el espectro promediado de larga duración, la modulación espectral, la semejanza de las voces que se mezclan, etc. Bronkhorst ofrece una exhaustiva revisión al respecto y concluye

que la inteligibilidad depende de varios factores que se relacionan de forma compleja, como son “las propiedades que las señales que interfieren, el número de señales, la configuración espacial de las fuentes y el entorno acústico” (Bronkhorst, 2000: 117).

Otra variable muy importante en la percepción del habla en ruido es la relación con la lengua en la que se presenta el estímulo (lengua materna *vs.* segunda lengua): contamos con numerosas investigaciones (iniciadas en los años 70: Lane, 1963; Gat y Keith, 1978, entre otros), donde se pone de manifiesto que el nivel de competencia en una segunda lengua influye de manera decisiva en la capacidad para identificar la señal de habla en ruido: incluso entre hablantes bilingües, si adquirieron la segunda lengua después de los 14 años (bilingües tardíos), necesitan más intensidad en la señal para poder diferenciarla del entorno e identificarla que los monolingües o bilingües tempranos (Mayo, Florentine y Buus, 1997); y cuanto más adversas sean las condiciones (más cercanas la intensidad del ruido y la de la señal), mayores serán las diferencias entre hablantes nativos y no nativos, con las correspondientes repercusiones para el rendimiento escolar de los niños (Crandell y Smaldino, 1996) y para su evaluación auditiva, incluso en el caso de los bilingües (Carlo, 2009). Hasta tal punto es persistente este fenómeno que algunos modelos predictivos de la inteligibilidad verbal incorporan el “factor no-nativo” para mejorar sus resultados (van Wijngaarden, Steeneken y Houtgast, 2002). Los efectos del ruido (estacionario, de habla en competencia o multihablante) sobre una señal lingüística en inglés por parte de aprendices españoles han sido estudiados en García Lecumberri y Cooke (2006) y Cooke, García Lecumberri y Barker (2008); estos investigadores analizan también qué parte de las dificultades se pueden atribuir al enmascaramiento energético y cuáles al informativo; según sus resultados, aunque en silencio la actuación de ambos grupos era similar, los no-nativos sufrieron mucho más que los nativos el efecto de cualquiera de los ruidos, siendo el multihablante el que más dificultades les supuso.

Los factores que influyen en las dificultades de los no-nativos para identificar la señal de habla en ruido son múltiples: la mayor o menor cercanía entre la L1 y la L2, la experiencia continuada en el uso de esta, la edad de adquisición, etc. En consecuencia, se han aducido razones diversas para explicar el fenómeno (Bradlow

y Alexander, 2007), que van desde el nivel segmental (la menor experiencia de los no-nativos les impide detectar las claves acústicas secundarias, aquellas que en condiciones normales no caracterizan a un sonido, pero en condiciones adversas sí asumen un papel primordial, porque resisten mejor al ruido), al prosódico (el ruido afecta a la percepción de cambios en la frecuencia fundamental que los nativos compensan por medio del conocimiento sintáctico, semántico y pragmático), o al semántico y contextual (las diferencias en predictibilidad de los contenidos en función del contexto facilita o dificulta la decodificación).

1.2.5. La variable edad

Los experimentos realizados con población infantil muestran que el efecto del ruido multihablante para la identificación del mensaje lingüístico es mayor en niños pequeños que en mayores. Las razones que se han dado para explicar este fenómeno se relacionan, por un lado, con la inmadurez en los sistemas de procesamiento, y por otro, con posibles diferencias en la estrategia de decodificación. Así, Fallon (2001) considera que los niños de cinco años, a diferencia de sus compañeros de nueve, parecen aplicar mecanismos globales para la identificación de pares mínimos, y no analíticos, lo cual afectaba especialmente a las palabras que solo se distinguían por un fonema situado en posición de coda silábica (*cap* vs. *cat*) presentadas en ruido. Sin embargo, todos los sujetos, niños y adultos, demostraron capacidad para utilizar la información en inicio de palabra, o la procedente del contexto, con el fin de mejorar su identificación. Incluso bebés de 24 meses mostraron en condiciones difíciles “habilidades suficientes para comprender el habla y mirar al referente apropiado”, cuando la diferencia de intensidad entre la señal y el ruido era solo de 5 dB (Newman, 2011), sin que el nivel de desarrollo léxico del niño resultara una variable determinante.

Otra línea de investigación sobre la percepción infantil del habla en ruido tiene relación con el ámbito clínico. Así, Ziegler *et al.* (2009) analizan las respuestas de niños con patologías del habla y dislexia, respectivamente; según sus resultados, en ambos casos la capacidad de discriminación está significativamente por debajo

de la que presenta la población normal en igualdad de condiciones (similar nivel de desarrollo lector), concluyendo que “la percepción de habla en ruido predice las habilidades lectoras subyacente a nivel de audición periférica (precolear), memoria, producción o habilidades atencionales” (Ziegler *et al.*, 2009: 742). También los niños con problemas de aprendizaje muestran especiales dificultades para interpretar el habla en ruido; en un estudio con técnicas de neuroimagen, un grupo multidisciplinar de la Northwestern University, en Chicago, encontró “anomalías en la representación sensorial fundamental de los sonidos, tanto a nivel cortical como en el tronco cerebral, en niños con problemas de aprendizaje cuando los sonidos del habla se presentaban en ruido, pero no en silencio” (Cunningham *et al.*, 2001: 758).

En cuanto al otro extremo de la variable edad, la vejez, se ha demostrado que incluso en personas que mantienen intacta su capacidad auditiva general, a medida que aumenta la edad se incrementan las dificultades para entender el habla en condiciones de ruido conversacional (Committee on Hearing, Bioacoustics and Biomechanics, 1988; Rajan y Cainer, 2008). En sujetos presbiacúscos (es decir, con una pérdida de audición asociada a la edad), el deterioro auditivo general explicaría un 66% de los resultados, pero el 33% restante tendría su origen en “un decremento general de la actuación debido a una eficiencia mental reducida, indicada por un enlentecimiento general y una menor capacidad de memoria” (van Rooij y Plomp, 1990: 2611). Estudios recientes con neuroimagen muestran, cuando las personas de más edad intentan identificar la señal de habla en ruido multihablante, una disminución de la activación en la corteza auditiva acompañada por un aumento de la actividad cortical en regiones cognitivas generales, probablemente como estrategia de compensación (Wong *et al.*, 2009).

1.2.6. La percepción del habla en ruido en la deficiencia auditiva

Afrontar las dificultades de las personas que sufren una pérdida de audición para discriminar el contenido del mensaje lingüístico en entornos ruidosos es uno de los retos de la Audiología desde hace varias décadas. Las sufren tanto quienes padecen pres-

biacusia, a quienes ya hemos aludido, como el resto de afectados por una disfunción auditiva, ya sea de oído medio, oído interno, de las vías auditivas o incluso del procesamiento auditivo central.

Las dos pruebas audiométricas funcionales más importantes son la audiometría tonal y la audiometría verbal; veremos muy rápidamente la diferencia entre ambas. Los tonos puros, sonidos formados por una sola onda y por tanto con una única frecuencia, son siempre estímulos artificiales, que tradicionalmente se producían mediante diapasones, y en el momento actual se generan electrónicamente. Los sonidos del habla, como todos los naturales, son ondas complejas, constituidos por la superposición de múltiples ondas simples. La estimulación mediante tonos puros se emplea en audiología, para evaluar la respuesta del sistema auditivo a frecuencias muy determinadas, variando las intensidades; este método, conocido como *audiometría tonal*, es la más común y generalizada de las pruebas subjetivas: es fácil de aplicar y proporciona una información fiable y estable. Sin embargo, desde hace muchos años se viene señalando (referencias en Cárdenas y Marrero, 1994; cf. también Taylor, 2003; Fallon, 2001) que el verdadero objetivo de una valoración auditiva debería ser cómo se perciben los estímulos que rodean a las personas en su entorno natural, y muy especialmente, cómo se descodifica el habla. Las diferencias perceptivas entre tonos puros y estímulos lingüísticos son consecuencia de sus diferentes características acústicas, pero también de su función: la señal lingüística tiene contenido y función comunicativa, es relevante biológicamente para nuestra especie⁴. Por eso no solo su proceso de descodificación es particular (incluso entre estímulos de habla se observan diferencias importantes si se trata de dígitos o de palabras: McArdle, Wilson y Burks, 2005), sino también su rendimiento en tareas de entrenamiento perceptivo y aprendizaje⁵.

⁴ Autores como Gentner consideran que la percepción de tonos puros es diferente a la del contenido comunicativo; según el autor, este hecho ya sucede en los animales que utilizan significaciones funcionales para comunicarse, como los pájaros de canto. En el estudio sobre la segregación del flujo acústico que vincula a la capacidad de localización espacial, afirma: “los parámetros acústicos relacionados con la segregación acústica de las señales comunicativas varían dramáticamente de la segregación de secuencias de tonos puros en animales” (Gentner y Ball, 2006: 658).

⁵ Millward *et al.* (2011) encontraron que los niños de 8-10 años mejoran sus resultados cuando se entrenan con palabras en ruido mucho más que con tonos puros.

La presencia de ruido afecta de forma muy particular a la comprensión del mensaje verbal; en consecuencia, los nuevos modelos de prótesis auditiva cuentan con mecanismos para minimizar en lo posible estas alteraciones, y en los últimos años se han ido desarrollando pruebas para evaluar la discriminación del mensaje oral (frases, palabras o dígitos) en presencia de ruido enmascarante (Killion *et al.*, 2006; Wilson *et al.*, 2007; McArdle y Wilson, 2008; para el español, Marrero y Cárdenas, 2012); Las razones para incluirlas en la práctica clínica se resumen en los puntos siguientes (Taylor, 2003):

1. En primer lugar, permiten abordar las quejas de los pacientes por su incapacidad para comprender el habla en ambiente con ruido de fondo.
2. Por otra parte, los datos de estas pruebas guiarán las decisiones para seleccionar la ayuda más apropiada para el paciente.
3. Ofrecen expectativas realistas sobre el beneficio que aportará el audífono al paciente cuando escuche con ruido de fondo.

2. ELABORACIÓN DE RUIDOS MULTIHABLANTE EN ESPAÑOL CON FINES AUDIOLÓGICOS

Para evaluar, diagnosticar y rehabilitar la percepción del habla en situaciones de pérdida auditiva es esencial, por las razones expuestas, considerar qué ocurre cuando la situación de habla no es la ideal, no es la que tiene lugar en la cabina insonorizada de la consulta. Para ello resulta imprescindible valorar la capacidad de discriminación del mensaje lingüístico en entornos ruidosos, y especialmente ante el tipo de ruido que más ha demostrado afectar a la comunicación: el ruido de habla. Como hemos expuesto anteriormente (cf. 1.1.2. Tipos de ruido), existen diferentes métodos para crear estos sonidos enmascarantes: imitándolo artificialmente (ruido con forma de habla), mediante superposición de grabaciones naturales (ruido multihablante), manipulando una grabación natural (habla inversa) o simplemente utilizando una locución de habla natural como ruido de competencia.

En el caso de la evaluación audiológica necesitamos mantener un delicado equilibrio entre el enmascaramiento energético y el informativo, procurando controlar algunas variables importantes en este último: no interesa que el contenido informativo del ruido compita con el de la señal, puesto que esta ha sido construida siguiendo una serie de criterios estrictos (equilibrio fónico y silábico, verificación de la familiaridad del léxico, etc.); pero sí es importante que el enmascaramiento resulte efectivo, para lo cual será conveniente que el ruido contenga emisiones de locutores con el mismo sexo que el de quien haya grabado la señal. Y lo que es más importante: tenemos que evitar en la mayor medida posible las variaciones de intensidad a lo largo del ruido, puesto que un “pico” inesperado podría enmascarar más una parte de la señal que otra, produciendo sesgos indeseados en los resultados.

En cuanto a la composición del ruido, es bastante habitual en el área de la evaluación audiológica enmascarar la señal con un ruido multihablante generado mezclando las voces de cuatro locutores (*four-talker babble*), especialmente desde que este tipo de máscara se incorporó a tests muy generalizados como el QuickSIN (Killion *et al.*, 2004; Wilson, 2003): estudios previos demuestran que la inteligibilidad de un mensaje se reduce un 40% por cada nuevo locutor que se incorpore a la grabación, hasta el tercero (Brungart, Ericson y Simpson, 2002), pero que con un número igual o inferior a tres el ruido se considera un parloteo o murmullo (*babble*), y solo a partir de cuatro su clasificación perceptiva adquiere carácter de “gentío” o “muchedumbre” (*large crowd*). La elección de cuatro locutores (que se considera perceptivamente equivalente a un punto intermedio entre el parloteo y el gentío, en una proporción de 66%-34%, según Krishnamurthy y Hansen, 2009) ofrece un buen equilibrio entre la complejidad de elaboración del ruido y las garantías de un resultado adecuado a nuestros fines.

A continuación se describirá la elaboración de un ruido multihablante de cuatro locutores en español con fines audiológicos; comenzaremos por exponer la metodología de elaboración y edición, se presentarán las características acústicas del sonido producido, en términos de espectro promediado de larga duración (*long-term averaged spectrum, LTAS*) y de variaciones de intensidad; por último se proporcionarán los resultados de una evaluación experimental del ruido y también sus efectos en la evaluación audio-

lógica. Además del ruido en habla adulta, adelantaremos también los resultados preliminares de un ruido de habla infantil que está en proceso de validación.

2.1. *Habla adulta*

2.1.1. Metodología: elaboración y edición

A) LOCUCIÓN

Partimos de una lista de 120 frases fónicamente equilibradas (Castro Bleda, 1998), para elaborar cuatro textos diferentes, que se asignaron aleatoriamente a los cuatro locutores, que tuvieron la posibilidad de familiarizarse con ellos antes de la grabación.

B) LOCUTORES

Considerando que la grabación que este ruido iba a enmascarar fue realizada por una locutora seleccionada por presentar un tono de voz grave, para la composición del ruido multihablante se seleccionaron tres mujeres y un hombre de edad, nivel sociocultural y características similares (universitarios, no fumadores, 24-31 años).

C) GRABACIÓN

La grabación se realizó en la cabina insonorizada del Laboratorio de Fonética del CSIC en el Centro de Ciencias Humanas y Sociales en Madrid⁶. El equipo de grabación empleado para este fin constó de una mesa de mezcla modelo Alesis Multimix 16 USB y un micrófono de mesa AKG C2000B. El software de grabación que se utilizó fue Adobe Audition 1.0. La grabación se realizó en mono, con una frecuencia de muestreo 44100Hz y resolución a 16 bits.

- GRABACIÓN GRUPAL. En primer lugar, se colocó a cada uno de los sujetos a un lado de la mesa de la cabina con el micrófono en el centro de la misma. La distancia entre el micrófono y la boca de cada sujeto era de 30 cm aproximadamente. Se facilitó a los sujetos unos tapones de oídos

⁶ La elaboración y validación del ruido multihablante aquí descrito fue presentada como uno de los trabajos de evaluación del Máster en Estudios Fónicos del CSIC, por uno de los autores de este trabajo, M. J. Rodríguez Cruz.

para evitar interferencias entre ellos mientras leían simultáneamente los textos respectivos, durante un minuto y medio de grabación continua.

- GRABACIONES INDIVIDUALES. Posteriormente grabamos a cada locutor individualmente, con la misma distancia aproximada entre el micrófono y la boca, y sin tapones de oídos. En la última grabación individual se pidió a una misma locutora que leyera los cuatro textos elaborados.
- EDICIÓN. Utilizamos el software libre Audacity 1.2.6 (<<http://audacity.sourceforge.net/>>) para igualar la duración de las pausas, sincronizar las locuciones y mezclar los ficheros de audio, generando cuatro prototipos de ruido:
 - Grabación grupal → ruido tipo I
 - Superposición de las grabaciones individuales → ruido tipo II
 - Superposición del ruido tipo I y el ruido tipo II → ruido tipo III
 - Superposición de los cuatro textos leídos por una sola locutora → ruido tipo IV

D) EVALUACIÓN PRE-EXPERIMENTAL

Para valorar la capacidad enmascarante de los cuatro prototipos de ruidos multihablante elaborados realizamos un pre-experimento que incluía dos tareas: una primera de inteligibilidad del ruido en sí, y otra posterior para valorar su capacidad de enmascaramiento sobre una señal de habla diferente.

2.1.1.1. *Inteligibilidad del ruido*

Once jueces sin pérdida auditiva reconocida, universitarios y pertenecientes a dos grupos de edad diferentes (cf. más adelante, *La variable edad*), uno mayor de 50 años y otro menor de 35, intentaron identificar el mayor número posible de palabras en los cuatro prototipos de ruido, que les fueron presentados a una intensidad cómoda mediante auriculares de alta calidad⁷. Los resultados se resumen en la tabla y gráficos siguientes:

⁷ Beyerdynamic DT131 estereofónicos, gama de frecuencias de 20-18.000Hz, impedancia 40 Ω, sensibilidad 96 dB, coeficiente de distorsión < 0,2%.

Media de palabras identificadas		Desviación estándar
Ruido tipo I (A)	74,5 = 9,8%	9,3
Ruido tipo II (B)	51,6 = 7,0%	6,7
Ruido tipo III (C)	23,8 = 3,2%	2,4
Ruido tipo IV (D)	23,4 = 3,1%	4,3

TABLA 1. *Inteligibilidad en los cuatro modelos de ruido multihablante*

Como vemos, el ruido obtenido en la grabación grupal (tipo I, A en el gráfico) resultó el más inteligible de todos, seguido por el obtenido mediante la superposición de las cuatro grabaciones individuales (tipo II = B). En cambio, el ruido resultante de mezclar los dos anteriores (tipo III = C) resultó mucho más difícil para los jueces, igual que el compuesto por cuatro grabaciones de la misma locutora (tipo IV = D). El análisis de varianza (ANOVA de una vía para cuatro muestras independientes) permite rechazar con un 99% de confianza la hipótesis nula ($p = 0.001$); al análisis *post-hoc* de comparaciones múltiples de Tukey indica que las diferencias entre el tipo I y el tipo II no fueron significativas, ni tampoco las diferencias entre el tipo III

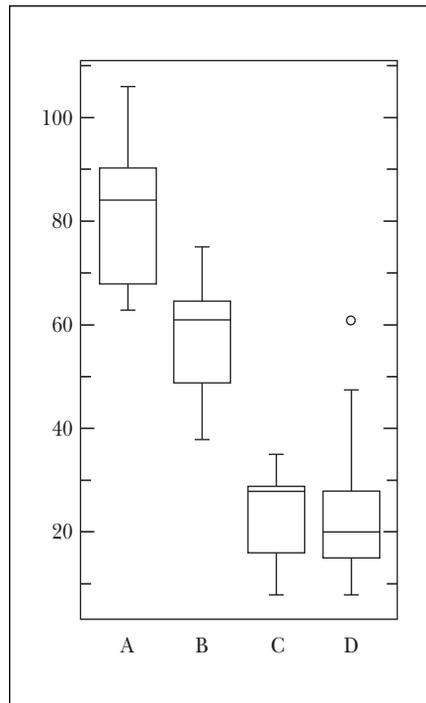


FIGURA 1. *Tasas de inteligibilidad en los cuatro modelos de ruido multihablante.*
 A = Ruido tipo I; B = Ruido tipo II;
 C = Ruido tipo III; D = Ruido tipo IV

y el IV, pero sí las de los dos primeros respecto a los dos últimos (I *vs.* III $p < .01$; I *vs.* IV $p < .01$; II *vs.* III $p < .05$; II *vs.* IV $p < .05$).

Los resultados de esta primera tarea pusieron de manifiesto, por lo tanto, que los prototipos más adecuados para nuestros fines eran el III y el IV, los menos inteligibles y los únicos en pasar a la fase siguiente.

2.1.1.2. *Discriminación de la señal enmascarada*

Para valorar la capacidad de enmascaramiento de los ruidos multihablante tipos III y IV se creó, además, una señal de habla a partir de unas listas de palabras fónicamente equilibradas elaboradas por Garrido y Llisterri para Widex y grabadas por un locutor masculino (Widex Audífonos, 1993⁸). Estas listas se sometieron a una tarea de discriminación por parte de los mismos sujetos que en la prueba anterior, tanto en silencio (condición de control) como con ruido superpuesto, con dos relaciones señal/ruido (S/R): 0 dB (la señal se presentó a la misma intensidad que el ruido) y -10 dB (la señal tenía 10 dB menos que el ruido).

Los resultados globales muestran que en la condición de control nuestros sujetos obtuvieron casi un 100% de aciertos, como era de esperar en normooyentes. Sin embargo, cuando la señal se presenta con ruido superpuesto aparecen los primeros errores (especialmente en relación señal/ruido de 0 dB) y sobre todo ausencia de respuesta, omisiones, que suponen el 50% de las respuestas cuando la intensidad del ruido y la de la señal son iguales, pero sube hasta el 97% cuando la primera es superior a la segunda.

Detectamos algunas diferencias entre el ruido III (superposición de los cuatro locutores en la grabación conjunta e individual) y el IV (superposición de los cuatro textos en una sola locutora): este último presentó menor capacidad de enmascaramiento, tanto en la condición S/R 0 dB (mayor tasa de errores y menor de omisiones que el ruido III) como en S/R -10 dB (más errores y menos omisiones); recordemos que el locutor de la señal era un hombre y la voz enmascarante era femenina, lo cual se ha señalado en la bibliografía como un factor que disminuye el efecto de la máscara.

⁸ Concretamente, utilizamos las listas n.º 3, 4 y 5 para las condiciones de enmascaramiento y las n.º 5 y 6 como control.

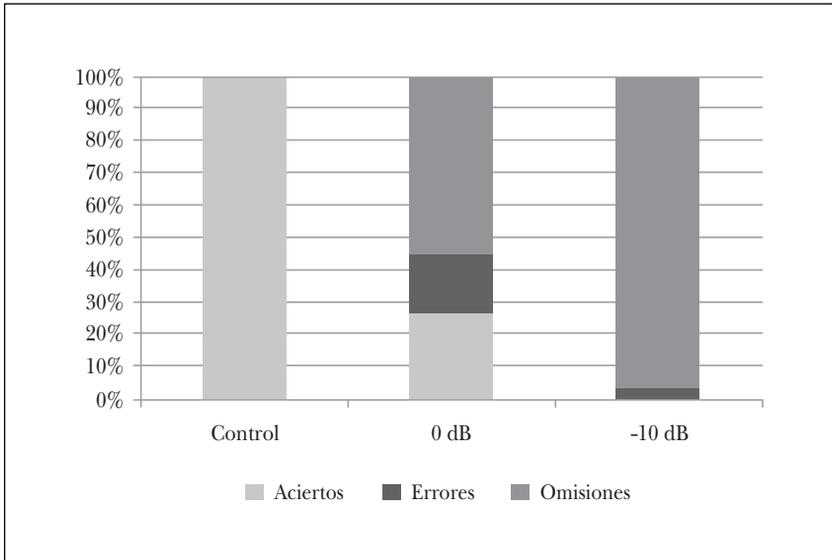


FIGURA 2. *Resultados globales de discriminación de la señal en ruido multihablante - adultos*

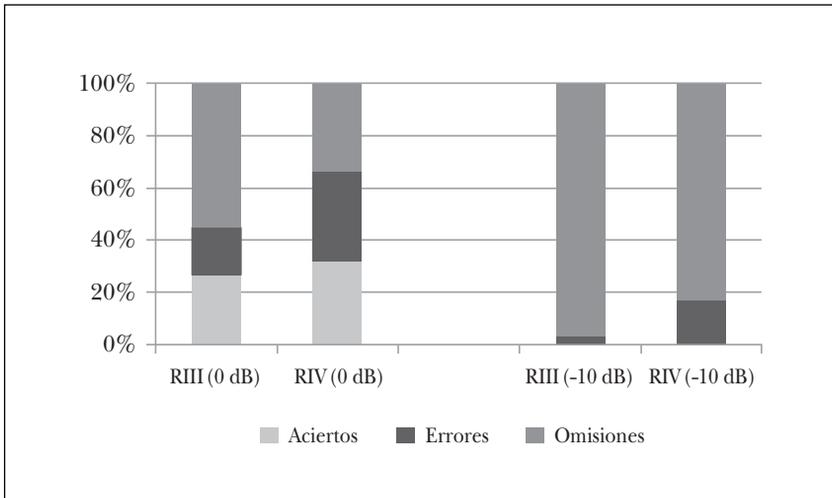


FIGURA 3. *Comparación de los ruidos tipo III y IV*

La prueba *t* de Student indicó que las únicas diferencias significativas ($p = 0,03$) se dieron entre errores y omisiones en 0 dB.

En definitiva, los resultados experimentales nos llevaron a considerar que el ruido más adecuado para utilizarlo en la elaboración de pruebas con fines audiológicos era el tipo III: cuatro locutores, tres mujeres y un hombre, leyendo cada uno un texto equilibrado, grabados de forma individual y grupal, y superpuestas las cuatro grabaciones individuales a la grupal. Y así se hizo, como veremos en 2.1.4.

2.1.1.3. *La variable edad*

Aunque no es el objeto de este trabajo, en la evaluación experimental separamos los jueces en dos grupos, uno de mayores de 50 años, y otro de menores de 35, con el fin de valorar las posibles repercusiones del proceso normal de envejecimiento auditivo en los resultados. Aunque, en general, los jóvenes obtuvieron mayores tasas de acierto que los participantes de más edad, las diferencias solo alcanzaron un nivel estadísticamente significativo en el ruido de tipo II (superposición de las grabaciones individuales). Por lo tanto, en esta pequeña muestra de normooyentes, la variable edad (joven/mayor) no resultó realmente relevante.

2.1.1.4. *Evaluación audiológica*

El objetivo final del ruido multihablante, además de su utilidad intrínseca como instrumento para enmascarar cualquier señal lingüística, fue utilizarlo para elaborar unas pruebas de audiometría verbal con frases equilibradas (fónica y sintácticamente) en ruido⁹. En Marrero y Cárdenas (2012) se sintetizan las características básicas de esa batería clínica; de forma muy resumida, se trata de un conjunto de 20 listas, cada lista está compuesta por seis

⁹ En el marco de un contrato de investigación Amplifon España-UNED. Se realizaron dos baterías: una locutada en silencio, y otra con ruido presentado mediante auriculares a la locutora, para provocar lo que se conoce como “efecto Lombard”, un conjunto de estrategias en la emisión del habla destinadas a mejorar la discriminabilidad del mensaje: mayor rango de frecuencias, más intensidad, menor tasa de habla y cambios rítmicos, entre otros (Folk y Schiel, 2011). En esta ocasión solo presentamos los resultados de las listas locutadas en silencio.

frases y cada frase contiene cinco palabras clave que el paciente o participante debe repetir exactamente¹⁰. Las frases se enmascararon con ruido a intensidad creciente, empezando por una relación señal/ruido de +15 dB (más intensa la señal que el ruido), y bajando en escalones de 3 dB: la segunda frase será 12 dB más intensa que el ruido; la tercera, 9 dB; la cuarta, 6 dB; la quinta, 3 dB; y la última presentará la misma intensidad que el ruido.

A continuación presentamos los resultados obtenidos en oídos normales y patológicos.

2.1.1.4.1. Evaluación audiológica en oídos normales

Las listas se han comenzado a validar en cinco centros de Amplifon en Madrid; hasta el momento contamos con 47 respuestas de normooyentes, acompañantes de los hipoacúsicos que acudían al centro a efectuar su revisión periódica; la edad media de este grupo es de 63.5 años, muy superior a la del grupo de evaluación

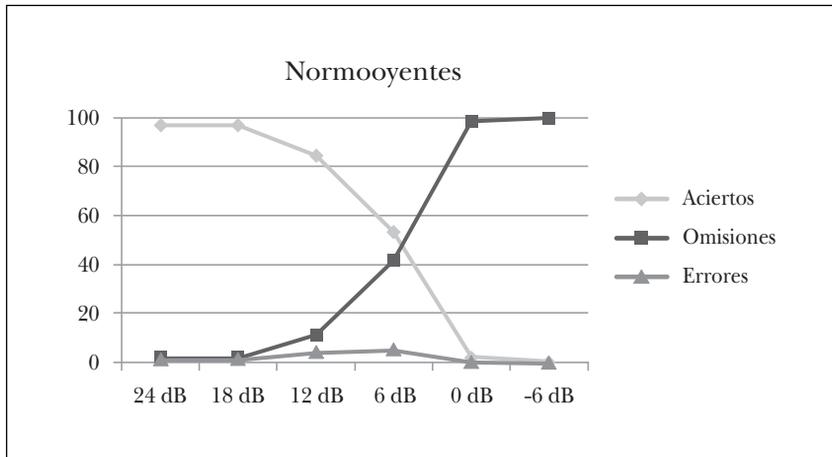


FIGURA 4. Discriminación verbal de frases equilibradas presentadas en ruido. Oídos normales. Adultos

¹⁰ Por ejemplo, una de las frases es “*El presidente del gobierno espera muchos cambios*”; en negrita las palabras clave que se deben repetir con exactitud.

pre-experimental. La presentación de los estímulos se realizó en la cabina audiométrica mediante campo libre, a través del audiómetro calibrado mediante un tono puro previo a 60 dB SPL.

El efecto del ruido multihablante sobre las frases fue más acusado en esta muestra que el producido sobre las palabras en la evaluación experimental: si entonces veíamos que a la misma intensidad que el ruido, la señal se conseguía identificar entre un 27% y un 32%, la tasa de identificación de frases en esta condición ofreció una tasa de aciertos del 0% (no se identificó ninguna palabra clave de la última frase), lo mismo que ocurría cuando la intensidad del ruido sobrepasaba la de la señal. Para alcanzar el 50% de aciertos necesitamos un incremento de 6 dB en la intensidad de la señal. Cuando este es de 12 dB se identifica correctamente un 85% de las palabras clave, y a partir de ahí la tasa de aciertos es del 97%.

Las diferencias entre esta muestra y la de normooyentes en la evaluación experimental posiblemente se explican atendiendo a la distinta composición sociocultural y sobre todo de edad entre ambas muestras; esta se ajusta más al perfil medio del usuario de prótesis auditiva.

2.1.1.4.2. Evaluación audiológica en oídos patológicos

En la población hipoacúsica es importante diferenciar al grupo que padece una disfunción de origen neurosensorial (en la cóclea o las vías auditivas) de quienes pierden audición por una alteración del oído medio (hipoacusia conductiva): la primera tiene efectos más graves en la capacidad para discriminar el habla, especialmente en entornos ruidosos. En nuestra muestra contamos con 32 valoraciones por hipoacúsicos neurosensoriales, 7 por conductivos y 8 por personas que se diagnosticaron como hipoacusia mixta (a la pérdida conductiva se une la neurosensorial); es necesario, por lo tanto, tomar con cautela los datos que se presentan a continuación, que son solo provisionales. Las condiciones de presentación fueron similares a las de los normooyentes.

Como vemos, los pacientes con pérdida neurosensorial tienen su techo de aciertos, en las condiciones más favorables (relación S/R de +24 y +18 dB), en un 72%; en el siguiente escalón la pérdida de discriminación es exponencial: a +12 dB de S/R apenas se supera el 40% de aciertos, y a +6 dB no se alcanza el 10%; a

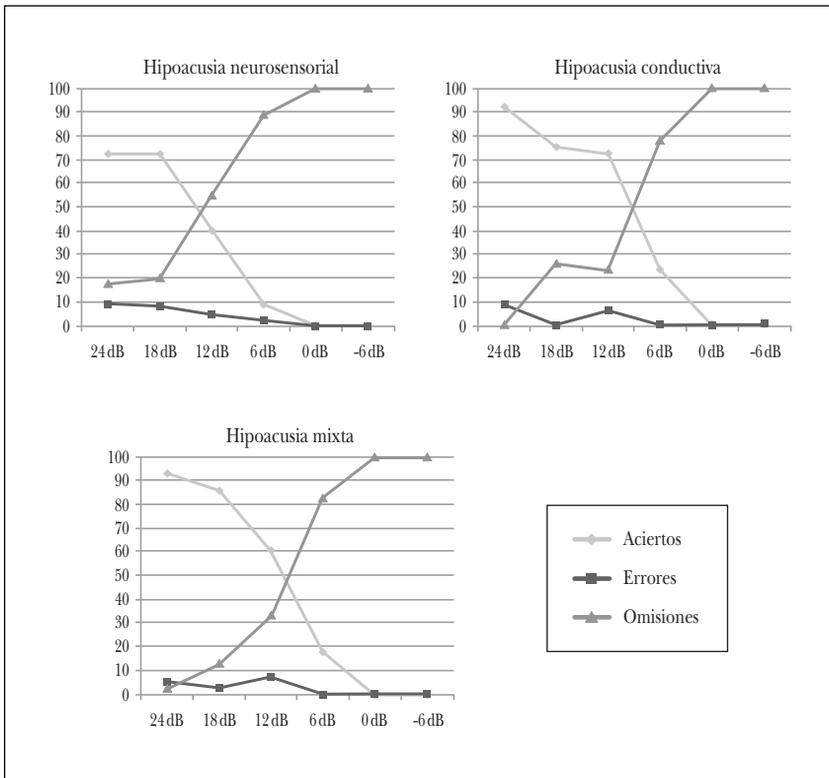


FIGURA 5. Discriminación verbal de frases equilibradas presentadas en ruido. Oídos patológicos. Adultos

partir de entonces no se obtiene ninguna respuesta. En cambio, las hipoacusias conductiva y mixta presentan un mejor patrón de respuesta: para empezar, unas tasas de acierto superiores al 90% en relaciones S/R de +24 dB; aunque en la segunda frase de cada lista, cuando disminuye esa diferencia en 6 dB, se registra una caída importante en la discriminación que no presentaban los normooyentes; y mientras aquellos conseguían mantener unas tasas de acierto superiores al 50% a S/R + 6 dB, estos apenas llegan al 18-23% en esa misma situación.

La diferencia entre los normooyentes y los tres grupos de hipoacúsicos resultó estadísticamente muy significativa en todos

los casos ($p < 0.0001$); sin embargo, las diferencias entre los tres diagnósticos no alcanzaron el nivel mínimo para considerarse significativas.

En definitiva, vemos cómo la aplicación del ruido multihablante descrito en este trabajo a una batería de pruebas logoaudio-métricas produce un resultado que, en esta primera etapa de validación, promete proporcionar unas curvas de respuesta claramente diferenciadas en oídos normales y en personas con pérdida auditiva; una vez realizada la adaptación protésica, en la medida en que la curva inicial se consiga acercar a la de los normooyentes, se estará consiguiendo el mejor rendimiento posible de la prótesis para la discriminación de la señal lingüística en situaciones de ruido de habla: un objetivo primordial para todo usuario de audífonos (Taylor, 2003).

2.1.2. Características acústicas

El espectro promediado de larga duración del ruido multihablante adulto en español, obtenido mediante el software Praat (<<http://www.fon.hum.uva.nl/praat/>>) tiene la siguiente distribución de energía (de 0 a 22 kHz); su intensidad media fue de 66,8 dB, con un máximo de 70,4 dB, un mínimo de 62,5 dB y desviación estándar = 1,28 dB:

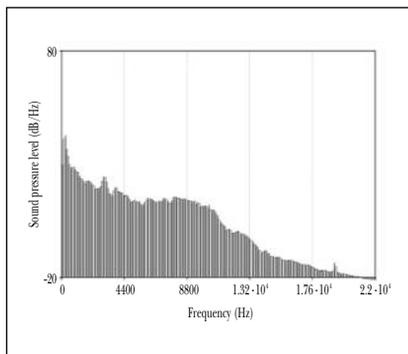


FIGURA 6. *LTAS del ruido Multihablante UNED-CSIC - adultos*

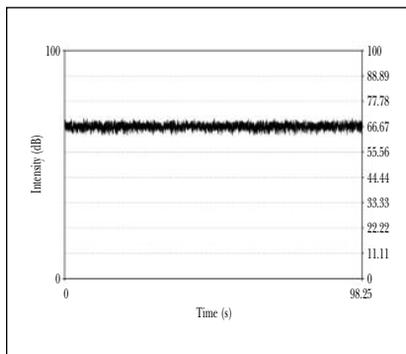


FIGURA 7. *Intensidad del ruido multihablante UNED-CSIC - adultos*

2.2. *Habla infantil*

Con unos objetivos similares a los planteados para el modelo adulto, elaboramos también un ruido multihablante con niños, que actualmente se encuentra en fase de evaluación, por lo que la información siguiente se ha de considerar como parte de un trabajo en curso.

2.2.1. Metodología: elaboración y edición

2.2.1.1. *Locución*

En esta ocasión partimos de un texto fónicamente equilibrado bien conocido en el ámbito de la fonética del español: el relato “El viento norte y el sol” (<<http://www.lengamer.org/publicaciones/ilusfon.php>>). Como locutores, para la composición del ruido multihablante infantil se seleccionaron dos niños y dos niñas de 6-7 años de edad.

2.2.1.2. *Grabación*

La grabación se realizó también en la cabina insonorizada del Laboratorio de Fonética del CSIC en el Centro de Ciencias Humanas y Sociales en Madrid, con el mismo equipo, y en esta ocasión en estéreo.

En primer lugar se realizó una grabación individual de cada niño leyendo el texto mencionado. Posteriormente se juntó a los dos niños y una de las niñas en una grabación grupal de habla espontánea, para la que contamos con tres láminas de dibujos (relacionadas con personajes de cuentos: *Los tres cerditos*, *Caperucita Rosa y Pocoyó*); la segunda niña fue grabada en una sesión posterior, y también realizó una lectura y una sesión de habla espontánea con las mismas láminas.

2.2.1.3. *Edición*

Utilizamos el software Adobe Audition para igualar la duración de las pausas, sincronizar las locuciones y mezclar los ficheros de audio, generando un primer prototipo mezclando las locuciones

individuales de los dos niños y las dos niñas, más las dos sesiones espontáneas superpuestas; la valoración por parte de tres expertos nos llevó a eliminar la sesión de lectura de uno de los dos niños, porque su destreza en lectura era incipiente e imprimía un tono demasiado monocorde al conjunto. Posteriormente se nivelaron los picos de intensidad.

2.2.2. Evaluación audiológica. Oídos normales

A la vista de los resultados obtenidos para adultos, y considerando la mayor sensibilidad auditiva de la población a la que nos dirigimos, niños de primaria (Trehub *et al.*, 1988), decidimos reducir el rango de enmascaramiento: la primera frase se presenta sin ruido perceptible (S/N 30 dB); la segunda es 10 dB más intensa que el ruido; la tercera, 5 dB; en la cuarta la intensidad de ambos es igual; en la quinta el ruido supera a la señal en 5 dB, y en la sexta en 10 (-10 dB S/N). Para comprobar si este patrón era adecuado, llevamos a cabo un pre-test en el cual se enmascararon las 60 frases de la prueba siguiendo esta pauta: las diez primeras sin ruido, las diez siguientes a 10 dB; las diez siguientes a 5 dB, y así sucesivamente. Diez niños normooyentes, de 7-8 años intentaron identificar las palabras clave de las frases, con el siguiente resultado:

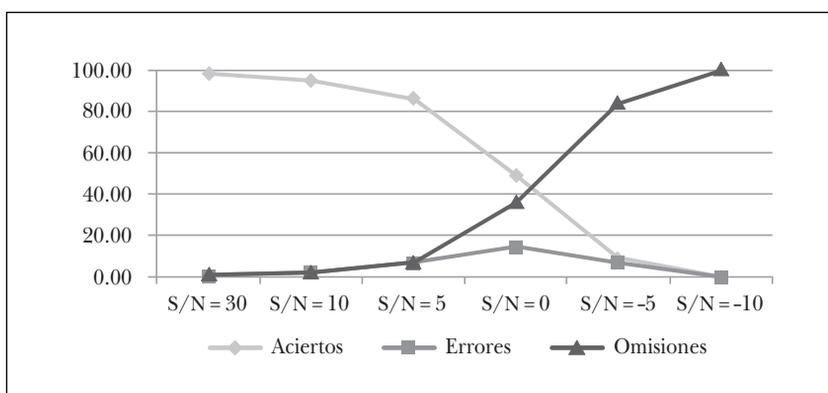


FIGURA 8. Discriminación verbal de frases equilibradas presentadas en ruido PIP-UNED. Oídos normales. Niños

Como vemos, el patrón general es muy similar al de los normooyentes adultos (objetivo que se trataba de conseguir): la tasa de aciertos es muy alta en las dos primeras condiciones, desciende en la tercera hasta un 86%, y en la cuarta se sitúa en un 50%; en la quinta apenas supera el 8%, y en la última es prácticamente nula. Sin embargo, si observamos los valores del eje horizontal, en niveles de enmascaramiento hay una diferencia importante, a favor de los niños: los adultos (o ancianos, en realidad) de nuestra muestra anterior necesitaron entre 5 y 7 dB más en la señal para conseguir las mismas tasas de identificación que los niños, en todas las condiciones.

2.2.3. Características acústicas

El espectro promediado de larga duración del ruido multihablante infantil, obtenido mediante el software Praat (<<http://www.fon.hum.uva.nl/praat/>>) tiene la siguiente distribución de energía (LTAS de 0 a 22 kHz). Si lo comparamos con el LTAS del ruido adulto comprobamos que su distribución es similar en frecuencias conversacionales (hasta 8-10 KHz), pero este tiene muchos más componentes en bandas superiores. La intensidad mínima fue de 85.23 dB y máxima de 89.2 dB; desviación estándar = 0.69 dB.

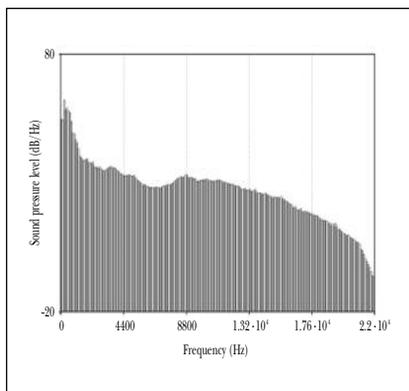


FIGURA 9. *LTAS del ruido multihablante UNED-PIP. Niños*

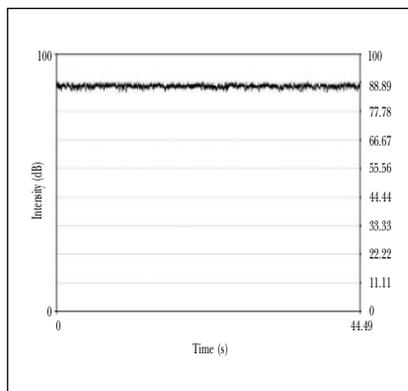


FIGURA 10. *Intensidad del ruido multihablante UNED-PIP. Niños*

3. DISCUSIÓN Y CONCLUSIONES

El habla es una señal muy compleja, no solo desde un punto de vista físico, sino también, y muy especialmente, desde un punto de vista cognitivo: como hemos visto en la introducción, son numerosas las variables que intervienen en su decodificación, especialmente en entornos ruidosos, un reto para el cual posiblemente contamos con un mecanismo neurofisiológico específico. Como consecuencia, las decisiones sobre cómo evaluar este proceso deben tomarse con las máximas cautelas, especialmente si esa evaluación tiene carácter clínico o rehabilitador y aspira a ofrecer cierta capacidad de generalización a las condiciones naturales en las que se desarrolla nuestra actividad comunicativa diaria. La composición adecuada del ruido de habla que vamos a superponer a la señal es, en este contexto, un elemento más, y ha de ser atendido con la misma atención que el resto.

En la bibliografía se han señalado las diferencias entre una grabación grupal (de personas conversando simultáneamente) y la superposición de varias grabaciones individuales (Krishnamurthy y Hansen, 2009); en nuestro caso, como se ha expuesto en 2.1.1 y 2.1.2, hemos considerado los dos métodos, y han sido los resultados de la evaluación pre-experimental los que nos han decantado por uno de ellos. En cuanto al número de locutores, ya se ha justificado (2.1) la elección de cuatro; además de los argumentos expuestos entonces, hemos de considerar también que cuando el número de locutores es inferior a esa cifra las diferencias de tono entre las diversas voces permiten separar flujos de habla (Divenyi, 2004), algo que en nuestro caso resultaría un inconveniente.

En algunos trabajos del ámbito audiológico español se ha utilizado un ruido multihablante comercial construido con lengua inglesa; sin embargo, cuanto más se profundiza en el estudio del enmascaramiento informativo más evidentes se hacen los inconvenientes de esa opción, y las pocas garantías que ofrece de que el efecto obtenido pueda generalizarse a situaciones de habla habituales para los sujetos de nuestro entorno. Hoen *et al.* (2007) ponen de manifiesto que no solo factores fónicos particulares de cada lengua (como la base de articulación, el predominio de sonidos anteriores o posteriores), o prosódicos (tipo de acento, patrones entonativos), sino incluso el contenido léxico y semántico de

los fragmentos de habla utilizados para la creación de ruido pueden alterar sus efectos. La influencia de la lengua en la que se construye el ruido multihablante sobre sus efectos como máscara de la señal lingüística han sido puestos de manifiesto para el contraste inglés-chino por van Engen y Bradlow (2007), en danés-sueco por Rhebergen *et al.* (2005), y en español-inglés (aunque solo para identificar consonantes inglesas) en el trabajo citado de García Lecumberri y Cooke (2006)¹¹. En algunos de estos trabajos se pone de manifiesto la interrelación entre el efecto de la lengua utilizada en el ruido y el número de locutores: cuando este es elevado (seis), los efectos del ruido son similares sea cual sea la lengua utilizada, debido a que la posibilidad de identificar componentes con significado es mucho menor que con un número de locutores más bajo (dos).

En cualquier caso, los estudios sobre la discriminación del habla en ruido están resultando cada vez más importantes, en primer lugar, para conocer los mecanismos básicos de la descodificación del lenguaje, que se han demostrado mucho más robustos en el cerebro humano que en los sistemas computacionales (Darwin, 2008). Esta capacidad de adaptación general a las condiciones adversas en el sistema de la comunicación en algunos casos se basa en la redundancia del sistema lingüístico y perceptivo y su tolerancia a la pérdida de información, especialmente cuando el ruido es estable; sin embargo, cuando el ruido proviene de una fuente sonora semejante a la de la señal, en nuestro caso el habla, son necesarias estrategias más complejas, que conciernen a la identificación de claves perceptivamente relevantes para identificar fonemas o unidades lingüísticas.

Pero no solo en el ámbito teórico, sino también desde diversas áreas aplicadas resulta de interés conocer los efectos del ruido de habla sobre la señal lingüística, y poder experimentar al respecto en investigaciones interlingüísticas. Además de sus obvias repercusiones en tecnologías del habla, para mejorar el funcionamiento de los sistemas de reconocimiento automático, también en otras áreas, como la enseñanza de segundas lenguas, por ejemplo, se han

¹¹ Cuestión diferente, pero no menos interesante, son las variaciones interlingüísticas e interculturales en cuanto a la tolerancia al ruido y sus efectos, cf. Namba, Kuwano y Schick (1986) para una comparación entre alemanes y japoneses.

mencionado las diferencias entre los efectos del ruido respecto a la lengua materna (cf. 1.2.4). En el ámbito de los estudios sobre patologías del habla y del lenguaje, aparte de las evidentes implicaciones en los casos de pérdida auditiva (1.2.6), se ha señalado la capacidad de identificar el mensaje lingüístico en entornos ruidosos como un predictor de la capacidad lectora, la memoria, o la capacidad de atención (Ziegler *et al.*, 2009), cuyas bases, según detectan estudios sobre niños con dificultades de aprendizaje o en personas mayores, pueden tener un sustrato neurológico (Cunningham *et al.*, 2001; Wong *et al.*, 2009).

El ruido descrito en este trabajo está a disposición de los investigadores de cualquiera de estas áreas que lo requieran a los autores, para su uso con fines científicos y no comerciales.

REFERENCIAS BIBLIOGRÁFICAS

- AMERICAN STANDARDS ASSOCIATION (1960): *Acoustical Terminology SI*, 1-1960, Nueva York, American Standards Association.
- BARUTCHU, A. *et al.* (2010): "Audiovisual integration in noise by children and adults", *Journal of Experimental Child Psychology* 105 (1-2), 38-50.
- BERGLUND, B.; P. HASSMÉN y R. F. S. JOB (1996): "Sources and effects of low-frequency noise", *J. Acoust. Soc. Am.* 99 (5), 2985-3002.
- BRADLOW, A. R. y J. A. ALEXANDER (2007): "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners", *J. Acoust. Soc. Am.* 121 (4), 2339-2349.
- BRONKHORST, A. (2000): "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions", *Acustica* 86, 117-128.
- BRUNGART, D. S. (2001): "Informational and energetic masking effects in the perception of two simultaneous talkers", *J. Acoust. Soc. Am.* 109 (3), March 2001, 1101-1109.
- BRUNGART, D. S., M. A. ERICSON y B. D. SIMPSON (2002): "Design Considerations For Improving the Effectiveness of Multitalker Speech Displays", *Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, July 2-5, 2002* ICAD02-07.
- BRUNGART D. S., B. D. SIMPSON, M. A. ERICSON y K. R. SCOTT (2001): "Informational and energetic masking effects in the perception of multiple simultaneous talkers", *J. Acoust. Soc. Am.* 110, 2527-2538.
- CÁRDENAS, M. R. y V. MARRERO (1994): *Cuaderno de Logaudiometría*, Madrid, UNED.

- CARLO, A. (2009): "A review of the effects of bilingualism on speech recognition performance", en *Perspectives on hearing and hearing disorders: Research and diagnosis* 13, 14-20.
- CASTRO BLEDA, M. J. (1998): "Modelado acústico de unidades subléxicas mediante una aproximación basada en métodos estructurales-conexionistas", tesis Universidad Politécnica de Valencia. Consultado el 10-02-10 en <<http://www.dsic.upv.es/docs/bib-dig/tesis/etd-10272003-000915/PhThesisCastro98.pdf>>.
- COMMITTEE ON HEARING, BIOACOUSTICS AND BIOMECHANICS (CHABA) (1988): "Speech understanding and aging", *J. Acoust. Soc. Am.* 83, 859-820.
- COOKE, M., M. L. GARCÍA LECUMBERRI y J. BARKER (2008): "The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception", *J. Acoust. Soc. Am.* 123 (1), 414-427.
- CRANDELL, C. y J. SMALDINO (1996): "Sound field amplification in the classroom: Applied and theoretical issues", en F. Bess, J. Gravel y A. Tharpe (eds.), *Amplification for children with auditory deficits*, Nashville, TN, Bill Wilkerson Center Press, págs. 229-250.
- CUNNINGHAM, J., T. NICOL, S. G. ZECKER, A. BRADLOW y N. KRAUS (2001): "Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement", *Clinical Neurophysiology*, 112, 758-767.
- DARWIN, C. J. (2008): "Listening to speech in the presence of other sounds", *Philosophical Transactions of the Royal Society Biological Sciences*, 363, 1011-1021.
- DÉLANO, P., I. ROBLES y L. ROBLES (2005): "Sistema eferente auditivo", *Rev. Otorrinolaringol. Cir. Cabeza Cuello*, 65, 55-62. Consultado el 5/5/12 en <<http://www.sochiorl.cl/indices/pdfs/65-1/08.pdf>>.
- DIRKS, D. D. y D. R. BOWER (1969): "Masking Effects of Speech Competing Messages", *Journal of Speech and Hearing Research*, vol. 12, 229-245.
- DIVENYI, P. L. (2004): *Speech Segregation by Humans and Machines*, Kluwer Academic Publisher, Dordrecht, The Netherlands.
- FALLON, M. (2001) "Children's perception of speech in noise", tesis doctoral del Departamento de Psicología de la Universidad de Toronto.
- FOLK, L. y F. SCHIEL (2011): "The Lombard Effect in Spontaneous Dialog Speech", en *INTERSPEECH-2011*, 2701-2704. Consultado el 12/03/2012 en <http://www.isca-speech.org/archive/archive_papers/interspeech_2011/i11_2701.pdf>.
- FREYMAN, R. L., U. BALAKRISHNAN y K. S. HELFER (2001): "Spatial release from informational masking in speech recognition", *J. Acoust. Soc. Am.* 109 (5), Pt. 1, May 2001, 2112-2122.
- GARCÍA LECUMBERRI, M. L. y M. COOKE (2006): "Effect of masker type on native and non-native consonant perception in noise", *J. Acoust. Soc. Am.* 119, 2445-2454.

- GAT, I. B. y R. W. KEITH (1978): "An effect of linguistic experience. Auditory word discrimination by native and non-native speakers of English", *Audiology*, 17, 339-345.
- GENTNER, T. y F. BALL (2006): "A neuroethological perspective", en D. Pisoni y R. Remez (eds.), *The Handbook of Speech Perception*, Blackwell Publishing, págs. 653-675.
- GIRAUD, A. L. *et al.* (1997): "Auditory efferents involved in speech-in-noise intelligibility", *Neuroreport* 8 (7), 1779-1783.
- HOEN, M. *et al.* (2007): "Phonetic and lexical interferences in informational masking during speech-in-speech comprehension", *Speech Communication* 49, 905-916.
- HALL, J. W., J. H. GROSE, E. BUSS y M. B. DEV (2002): "Spondee recognition in a two-talker masker and a speech shaped noise masker in adults and children", en *Ear & Hearing* 23, 159-165.
- KAWASE, T., B. DELGUTTE y M. C. LIBERMAN (1993a): "Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones", *Journal of Neurophysiology* 70(6), 2533-2549.
- KAWASE, T. y M. C. LIBERMAN (1993b): "Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones", *Journal of Neurophysiology* 70(6), 2519-2532.
- KILLION, M. C., P. A. NIQUETTE, G. I. GUDMUNDSEN, L. J. REVIT y S. BANERJEE (2006): "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners", *J. Acoust. Soc. Am.* 119 (3).
- KILLION, M. C. *et al.* (2004): "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners", *J. Acoust. Soc. Am.* 116 (4), 2395-2405.
- KRISHNAMURTHY, N. y J. H. L. HANSEN (2009): "Babble Noise: Modeling, Analysis, and Applications", *IEEE Transactions on Audio, Speech, and Language Processing* 17 (7), 1394-1407.
- KUMAR, U. A. y C. S. VANAJA (2004): "Functioning of Olivocochlear Bundle and Speech Perception in Noise", *Ear & Hearing* 25 (2), 142-146.
- LANE, H. (1963): "Foreign accent and speech distortion", *J. Acoust. Soc. Am.* 35 (4), 451-453.
- MARRERO AGUIAR, V. (2008): "La fonética perceptiva: Trascendencia lingüística de mecanismos neuropsicofisiológicos", *Estudios de Fonética Experimental*, XVII, 207-245.
- MARRERO AGUIAR, V. (1990): "Perturbaciones de la percepción desde el punto de vista fonológico y fonético", Comunicación presentada en el I Simposio Mundial de Audiología, Madrid.
- MARRERO AGUIAR, V. y M. R. CÁRDENAS DÍAZ DE ESPADA (2012): "Audiometría verbal", en E. Salesa y J. Perelló (eds.), *Manual de Audiología*, Barcelona, Masson.

- MAYO, L. H., M. FLORENTINE y S. BUUS (1997): "Age of second-language acquisition and perception of speech in noise", *J. Speech Lang. Hear. Res.* 40, 686-693.
- MCARDLE, R. A., R. H. WILSON y C. A. BURKS (2005): "Speech recognition in multitalker babble using digits, words and sentences", en *Journal of the American Academy of Audiology* 16, 726-739.
- MCARDLE, R. y R. WILSON (2008): "Selecting Speech Tests to Measure Auditory Function", *ASHA Leader* 13(12), 5-6.
- MILLWARD, K. E. *et al.* (2011): "Training speech-in-noise perception in mainstream school children", en *International Journal of Pediatric Otorhinolaryngology* 75, 1408-1417.
- MORA, E., H. MARTÍNEZ y A. BETANCOURT (2007): "Los ruidos en nuestro entorno", *Lengua y Habla* n.º 11, págs. 57-69.
- MUCHNIK, C. *et al.* (2004): "Reduced Medial Olivocochlear Bundle System Function in Children with Auditory Processing Disorders", *Audiology and Neuro-Otology* 9, 107-114.
- NAMBA, S., S. KUWANO y A. SCHICK (1986): "A cross-cultural study on noise problems", *J. Acoust. Soc. Jpn.* 7(5), 279-288.
- NATIONAL ACADEMY OF ENGINEERING (2010): *Technology for a Quieter America*, Washington, The National Academies Press.
- NEWMAN, R. S. (2011): "2-year-olds' speech understanding in multitalker environments", *The Official Journal of the International Society of Infant Studies* 16, 447-470.
- PARIKH, G. y P. C. LOIZOU (2005): "The influence of noise on vowel and consonant cues", *J. Acoust. Soc. Am.* 118, 3874-3888.
- PRIETO, P., C. PUGLIESI, J. BORRÁS-COMES, E. ARROYO y J. BLAT (2011): "Cross-modal prosodic and gestural contribution to the perception of contrastive focus", en *12th Annual Conference of the International Speech Communication Association*, Florence, Italy, 2011. Consultado el 12/03/2012 en <<http://prosodia.upf.edu/membres/joanborras/linguistica/arxiu/publicacions/2011/interspeech.pdf>>.
- RAJAN, R. y K. E. CAINER (2008): "Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers", *Neuroscience* 154, 784-495.
- RHEBERGEN, K. S., N. J. VERSFELD y W. A. DRESCHLER (2005): "Release from informational masking by time reversal of native and non-native interfering speech", *J. Acoust. Soc. Am.* 118, 1274-1277.
- ROSE, F. (2002): *Forensic Speaker Identification*, Londres, Taylor and Francis.
- SCHNEIDER, B. A.; L. LI y M. DANEMAN (2007): "How competing speech interferences with speech comprehension in everyday listening situations", en *Journal of the American Academy of Audiology* 18, 478-591.
- TAYLOR, B. (2003): "Speech-in-noise tests. How and why to include them in your basic test battery", *The Hearing Journal* 56 (1), 40-44.

- TILLMAN, T. W, R. CARHART y W. O. OLSEN (1970): "Hearing Aid Efficiency in a Competing Speech Situation", *Journal of Speech and Hearing Research*, vol. 13, 789-781.
- TREHUB, S. E., B. A. SCHNEIDER, B. A. MORRONGIELLO y L. A. THORPE (1988): "Auditory Sensitivity in School-Age Children", *Journal of Experimental Child Psychology* 46, 273-285.
- VAN ENGEN, K. J. y A. R. BRADLOW (2007): "Sentence recognition in native- and foreign-language multi-talker background noise", *J. Acoust. Soc. Am.* 121 (1), 519-526.
- VAN ROOIJ, J. C. G. M. y R. PLOMP (1990): "Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses", *J. Acoust. Soc. Am.* 88 (6), 2611-2624.
- VAN WIJNGAARDEN, S. J., H. J. M. STEENEKEN y T. HOUTGAST (2002): "Quantifying the intelligibility of speech in noise for non-native listeners", *J. Acoust. Soc. Am.* 111, 1906-1916.
- WIDEX AUDÍFONOS, S. A. (1993): Audiometría vocal. Prueba de inteligibilidad. Discriminación acústica. Confusión consonántica. Listas elaboradas por el Departamento de Filología Española de la UAB- J. M. Garrido y J. Llisterri. Medida del poder separador temporal Test Leman-Renard. Elaborado por el Dr. Leman y X. Renard. Locator: J. Royo. CD Digital Audio.
- WILSON, R. H., R. A. MCARDLE y S. L. SMITH (2007): "An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss", *Journal of Speech, Language, and Hearing Research*, 50, 844-856.
- WILSON, R. H. (2003): "Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance", *Journal of the American Academy of Audiology*, 14 (9), 453-470.
- WONG, P. C. M., J. X. JIN, G. M. GUNASEKERA, R. ABEL, E. R. LEE y S. DHAR (2009): "Aging and Cortical Mechanisms of Speech Perception in Noise", *Neuropsychologia* 47(3), 693-703.
- ZIEGLER, J. C., C. PECH-GEORGE, F. GEORGE y C. LORENZI (2009): "Speech-perception-in-noise deficits in dyslexia", *Developmental Science* 12, 732-745.