

Rosario, Argentina, 12 y 19 al 23 de mayo de 2014

Ruido y contenido semántico

Federico Miyara^a, Vivian Pasch^{b,c}, Susana Cabanellas^b, Marta Yanitelli^b

^aLaboratorio de Acústica y Electroacústica; FCEIA, UNR, Rosario, Argentina

E-mail: fmiyara@fceia.unr.edu.ar

^bDepartamento de Física, FCEIA, UNR, Rosario, Argentina

^cTaller de Física I, FAPyD, UNR, Rosario, Argentina

1 Introducción

Tradicionalmente, la investigación de los efectos no fisiológicos del ruido ha abordado la relación entre el nivel de presión sonora con ponderación A y alguna valoración cuantitativa del efecto. Los indicadores que se han utilizado más frecuentemente son el grado o magnitud del efecto (por ejemplo, el porcentaje de sílabas reconocidas correctamente, ver figura 1) o la proporción de individuos en los que aparece el efecto (por ejemplo, el porcentaje de personas altamente molestas, ver figura 2).

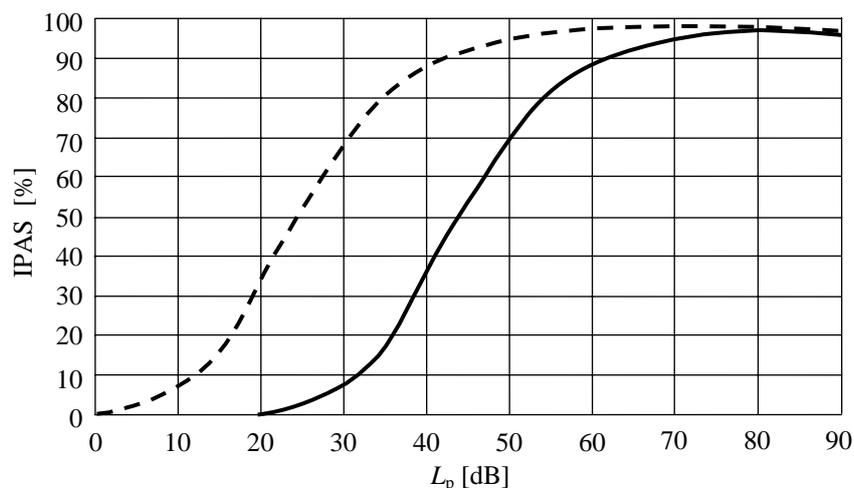


Figura 1. Índice porcentual de articulación silábica en función del nivel de presión sonora de la voz, en presencia de ruido de 43 dB de nivel de presión sonora. En línea de puntos se ha reproducido la curva medida sin ruido, con fines comparativos (Knudsen-Harris, 1978).

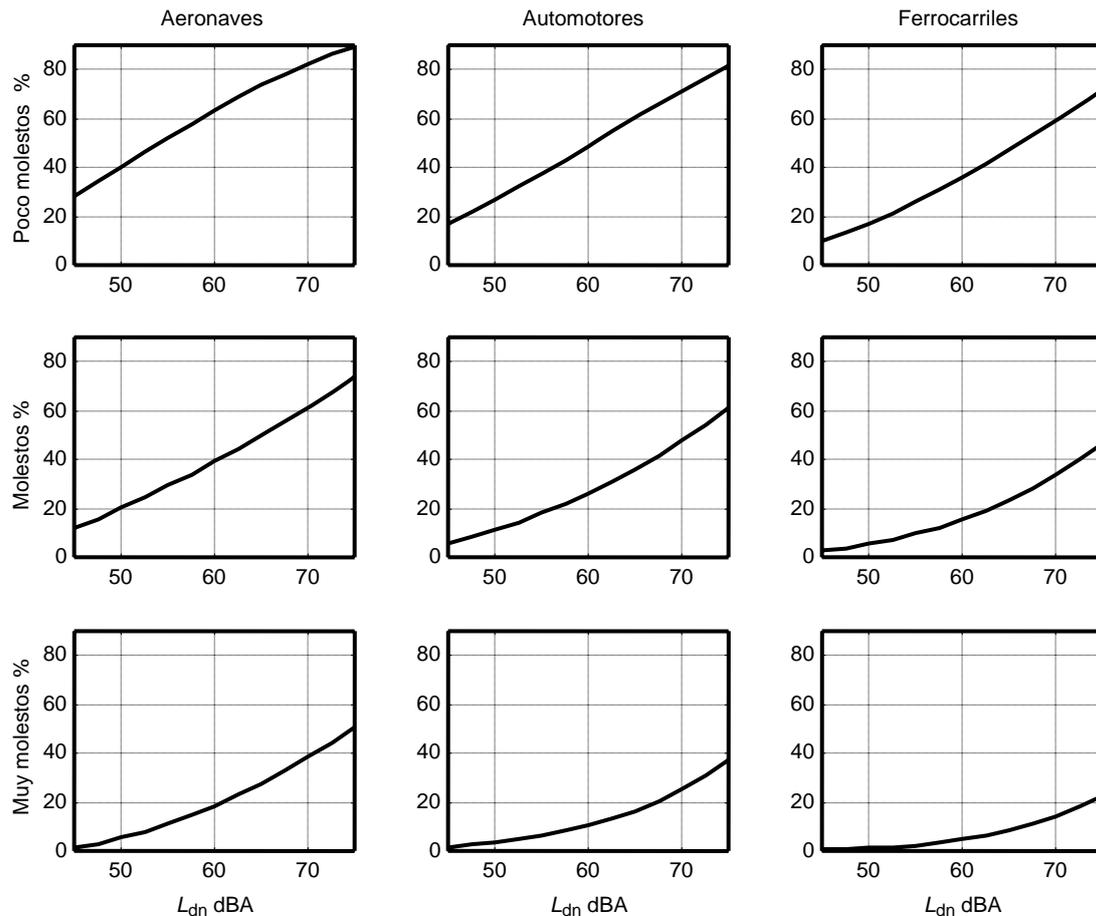


Figura 2. Curvas de Miedema para tres tipos de ruidos: tránsito aeronáutico, automotor y ferroviario. La serie de arriba da el porcentaje de personas poco molestas, la del centro el porcentaje de personas molestas y la de abajo el porcentaje de personas muy molestas en función del nivel día-noche L_{dn} (un promedio que penaliza los niveles nocturnos sumando 10 dBA a los valores medidos) [Miedema et al., 1998].

Dentro de los efectos no fisiológicos que merecen especial atención se pueden mencionar: molestia, incomodidad, intranquilidad, ansiedad, miedo, sobresalto, alarma, enmascaramiento, pérdida o disminución de la inteligibilidad de la palabra hablada, perturbación de la memorización a corto plazo, pérdida del alerta, perturbación del proceso de enseñanza y aprendizaje.

Muchos de estos efectos, particularmente la molestia, se correlacionan sólo débilmente con el nivel ponderado A, lo cual ha llevado a aplicar correcciones que tienen en cuenta diversas circunstancias, por ejemplo el horario, el carácter tonal o impulsivo, el contenido de baja frecuencia, la zona, etc. Esta circunstancia ha llevado a poner en tela de juicio el uso de dicho indicador.

Una característica de muchos tipos de ruido es la de poseer *contenido semántico*, es decir, contenido significativo. Sus efectos sobre el ser humano no se restringen, así, a los que se derivan de mecanismos puramente físicos o biológicos, como el fenómeno psicoacústico de enmascaramiento o la secreción de determinadas hormonas (adrenalina y noradrenalina), sino que desencadenan procesos cognitivos que interfieren o compiten con otros procesos simultáneos.

La capacidad de reaccionar cognitivamente ante estímulos con contenido semántico se vincula al estado de alerta. El alerta es una necesidad desde el punto de vista zoológico para conseguir el sustento o para enfrentar el peligro. Dicho estado implica la disposición para recibir y procesar información del entorno, información que muchas veces se manifiesta en forma de estímulos sensoriales. La vista, el oído y el olfato son los sentidos que permiten recibir información a distancia. De ellos el oído es el único que a la vez permite una percepción rápida, a distancias considerables y sin mediar contacto visual (es decir, aun en presencia de obstáculos), con buena localización direccional y espacial. De allí que la evolución haya favorecido un gran desarrollo de la percepción auditiva, que en el hombre llega al nivel de haber posibilitado el surgimiento del lenguaje, un sofisticado medio de comunicación simbólico.

El estado de alerta podría compararse con lo que en informática se denomina el sistema de *interrupciones*. Una interrupción es una señal que se presenta en forma aleatoria (es decir, en un instante imprevisible) y que desencadena un proceso diferente al que se estaba llevando a cabo. Según la situación, dicho proceso nuevo puede tener diversos niveles de prioridad, administrados en general por el sistema operativo. Una vez atendida la interrupción y ejecutado el proceso, el control vuelve al proceso que se estaba desarrollando previamente. Un ejemplo de ello es el oprimir una tecla en una computadora.

Este modelo sencillo ilustra la naturaleza de los fenómenos iniciales que tienen lugar en la mente humana ante una señal apropiada. La actividad que hasta ese momento acaparaba la atención es interrumpida y el interés se dirige ahora, voluntaria o involuntariamente, a la nueva señal, que deberá ser procesada a fin de determinar si es o no pertinente, si requiere o no una atención más prolongada, si desencadena otra serie de acciones o no. Dicho procesamiento puede requerir mayor o menor aporte de recursos mentales.

Existe una diferencia, sin embargo, con el caso informático. Los procesos mentales involucran recuerdos, asociaciones, intuiciones, analogías, inducciones, etc. Su interrupción no se reduce sólo a la suspensión momentánea de una secuencia lineal de acciones susceptible de ser retomada en cualquier momento sino, a menudo, de una compleja red que se construye en forma no unívoca, con intervención de elementos fortuitos e, inclusive, de percepciones subliminales. El regreso al punto de interrupción puede ser imposible. La red debe ser reconstruida a partir de algunos de los elementos originales, pero con intervención de elementos nuevos. Aun alcanzar un estado comparable o equivalente al original puede requerir un tiempo que se prolonga bastante más allá de la finalización del estímulo interruptor.

Este proceso de interrupción sería una de las causas de la sensación de molestia ocasionada por determinados tipos de ruido. Así, por ejemplo, a igual nivel y características espectrales, el ruido del tránsito resulta generalmente menos molesto que una música no deseada.

El objetivo de este trabajo es explorar y poner de manifiesto algunos mecanismos involucrados y algunas posibles estrategias de análisis para el estudio y la investigación sobre el contenido semántico del ruido y sus efectos.

2 Fundamentos conceptuales

El término “ruido” admite dos interpretaciones según se considere el punto de vista físico-objetivo o perceptivo-subjetivo. La primera se refiere al hecho de poseer un espectro frecuencial continuo, es decir que contiene todas las frecuencias audibles, en contraposición con los sonidos tonales o pseudotonales que poseen sólo un conjunto discreto de frecuencias constitutivas.

La segunda expresa simplemente que es “sonido no deseado”. No necesariamente el hecho de ser “no deseado” implica que se lo rechace o que automáticamente produzca efectos negativos. “No deseado” sólo implica que no se lo busca especialmente, que no hay intención de escucharlo por sí mismo. En otras palabras, puede ser no deseado y al mismo tiempo aceptárselo si se presenta espontáneamente. Pero, desde luego, la definición incluye también aquellos sonidos que por sus características (nivel, intermitencia, contenido semántico, etc.) provocan rechazo o causan efectos adversos.

Se considera que un sistema de comunicación posee un *emisor* que emite un *mensaje* materializado en una *señal*, un canal de transmisión, y un *receptor* que recibe la señal recuperando el mensaje [Fuentes, 1988]. En muchos casos, como en el de la radiofonía, el receptor no es único y puede ser indeterminado. En tales casos existe una intención de comunicar y un camino directo entre el emisor y cada receptor. En otros casos, como el de la música grabada, no hay un camino directo en sentido estricto, pero sí en sentido virtual.

Sin embargo, puede haber señales susceptibles de ser recibidas por un receptor que, sin haber sido emitidas con intención de comunicar, posean contenido semántico. Se denomina *indicio* a este tipo de señal. El ruido urbano es un ejemplo de ello: el ruido del tránsito, el de las maquinarias, el de la construcción, etc. Inclusive los denominados *signos naturales* [Marty et al., 1995], como el trueno o el canto de los pájaros, se encuadran en esta categoría. Los indicios son menos abstractos que otros tipos de signos, ya que no son arbitrarios sino el resultado de algún fenómeno físico.

Es posible y conveniente extrapolar algunas cualidades y caracteres de las señales a los indicios. Así, por ejemplo, muchos indicios poseen *carácter apelativo*, en el sentido de que atraen al receptor a prestarles atención. Esto no es en realidad una cualidad del indicio sino del receptor, su experiencia previa, su historia individual, etc.

Pero ¿qué es *contenido semántico*? Recurriendo al concepto de *triángulo semiótico* introducido por Ogden y Richards y modificado por Stern y Ullmann [Fuentes, 1998], formado por la terna *referente - significado - significante* (donde el referente es el objeto real, el significado su idea mental y el significante la imagen externa, en este caso acústica, empleada para representarlo), el contenido semántico es la existencia de información inteligible para el receptor, capaz de evocar la *representación* de un referente. En otras palabras, puede haber contenido semántico aun en un fenómeno sin intención comunicativa, en la medida en que sea capaz de provocar una representación como imagen mental (de cualquier tipo) en un receptor. Este proceso de construcción de la imagen mental se conoce como *semiosis* [Marty et al., 1995].

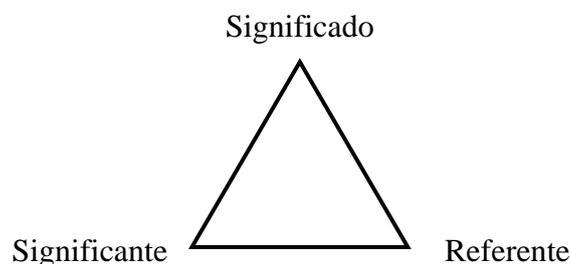


Figura 3. Triángulo semiótico de Ogden y Richards modificado por Stern y Ullmann [Fuentes, 1998].

A nuestros fines poco importa si la representación es correcta, ya que no existió realmente intención comunicativa. Tampoco interesa si existió un emisor consciente de que emitía un mensaje o, para el caso, si realmente se trató o no de un mensaje. Lo que interesa es el efecto distractivo que produce el proceso de semiosis de un fenómeno con cualidades de mensaje.

3 Contenido semántico

El contenido semántico puede ser directo, indirecto o evocado. Es *directo* cuando aparece explícitamente un significante, es decir, por ejemplo, una palabra o conjunto de palabras pronunciadas con suficiente nivel y relación señal/ruido como para ser inteligibles. En este caso, presumiblemente existió intención comunicativa en el emisor, aun cuando el destinatario del mensaje fuera un sujeto diferente de aquel que lo recibe como ruido, es decir, como sonido no deseado. Este tipo de situaciones se da frecuentemente en ámbitos públicos tales como oficinas, bares, restaurantes, aulas de clases e, inclusive, en bibliotecas. Probablemente sea éste uno de los casos que más atención ha recibido por parte de los investigadores, sobre todo por el inglés Dylan Jones y sus colaboradores, bajo la designación genérica de *sonido irrelevante* o *habla irrelevante* [Jones, 1995; Jones et al., 1995; Jones et al., 1990; Jones et al., 1992]. Otra situación típica se da en la exposición a música. Si bien el contenido semántico de la música no tiene el mismo alcance que en la palabra hablada, la existencia de estructuras rítmicas, melódicas y armónicas conocidas por el receptor acarrearán contenido significativo para éste.

El contenido semántico es *indirecto* cuando no aparece un significante explícito o aparece parcial o fragmentariamente. Se da esta situación cuando el mensaje se ve afectado por otro ruido (por ejemplo el efecto “cocktail party”), por enmascaramiento o por distorsión lineal (por ejemplo, la escucha a través de un tabique) o no lineal (por ejemplo, la escucha mediante un sistema de amplificación defectuoso). También cuando se reconoce que se trata de habla en el propio idioma, aun cuando no se reconozcan las palabras, o cuando se escucha una lengua extranjera desconocida o incomprensible.

El contenido semántico es *evocado* cuando, sin existir un estímulo normalmente asociado a un mensaje, se presentan ciertas condiciones de similitud que podrían desencadenar procesos mentales característicos de la decodificación de un mensaje. Puede haber similitud espectral (es decir, contenido de frecuencias semejante al de algún tipo de mensaje), similitud de los patrones temporales (por ejemplo, una evolución en el tiempo similar a la de la palabra hablada), o similitud de situaciones (una señal que sustituye a otra que normalmente está asociada a determinada situación).

La categoría de contenido semántico evocado es la más difícil de precisar y de estudiar, tanto en su caracterización fenomenológica como en la determinación precisa de sus mecanismos, algunos de los cuales pueden tener una fuerte base psicológica.

Es importante distinguir el concepto de contenido semántico del concepto de significado. Conceptualmente una señal posee contenido semántico si es capaz de desencadenar un proceso mental de semiosis o extracción de significado.

4 Experimentos sobre habla irrelevante

Un posible camino para estudiar el contenido semántico del ruido es examinar los efectos en comparación con los que provocan señales que poseen contenido semántico. Entre ellas se destaca la palabra hablada. A efectos de evitar considerar el caso en que se trata de la señal principal o el objeto específico de atención, es habitual trabajar con lo que se denomina habla irrelevante, es decir, señales vocales que compiten con otra actividad mental.

Los primeros experimentos sobre los efectos del habla irrelevante sobre tareas o procesos cognitivos fueron realizados por Colle y Welsh [Colle et al., 1976] (véase también [Jones et al., 1992]). En los mismos se examinó la interrupción de la memoria del orden serial. Se les mostraban a los sujetos 8 consonantes (por ejemplo F, K, L, M, Q, etc.), de a una por vez. Luego de 10 s se les pedía escribir las consonantes en el mismo orden en que las habían visto. A algunos sujetos se les pasaba una grabación en un idioma extranjero (alemán) con intensidad moderada y el resto trabajaba en silencio. Los sujetos expuestos al habla irrelevante cometían más errores que los que trabajaron en silencio. Se realizaron pruebas a dos niveles sustancialmente diferentes (48 dBA y 76 dBA), no obteniéndose diferencia significativa. En algunos casos se requería al sujeto realizar una resta durante el intervalo, obteniéndose el resultado notable de una disminución del efecto disruptivo del habla irrelevante. También se obtuvieron diferencias de performance en función de la similitud fonológica de las consonantes (por ejemplo, la rima entre sus nombres: “efe”, “ese”, “ene”, en contraposición con “ka”, “te”, “qu”) [Jones et al., 1995]. Las listas de consonantes diferentes eran más afectadas que las listas similares.

Según Jones y Morris [Jones et al., 1992], el habla irrelevante interfiere con tareas que involucran la memorización y transformación de patrones visuales, particularmente en el caso de material escrito. Al parecer el material escrito sufre una transformación mental de grafema a fonema en el que queda codificado según su representación fónica más que en su aspecto visual o, inclusive, su significado. Las características físicas como la dirección de procedencia o la intensidad no tienen efecto significativo en este experimento, así como tampoco las características semánticas del habla ya que se trataba de una lengua extranjera.

Jones realizó otros experimentos asignando a los sujetos la tarea de corregir un texto con errores [Jones et al., 1990]. El texto a corregir se mostraba en un monitor de computadora y había que detectar dos tipos de errores: *contextuales*, por ejemplo el tiempo de verbo o la concordancia de número, y *no contextuales*, por ejemplo los errores ortográficos. La detección de errores contextuales requería la comprensión del texto. Los errores no contextuales no requerían comprensión y podían detectarse en palabras aisladas.

La corrección de texto en sí involucra dos procesos: 1) Análisis de significado, que es la tendencia natural frente a un texto narrativo extenso, y 2) Detección de errores ortográficos, proceso menos automático. Son dos actividades incompatibles que requieren ser realizadas separadamente.

Para favorecer uno u otro proceso los investigadores permitían al sujeto visualizar cinco líneas de texto o sólo una. Desde el punto de vista cognitivo, implica dos diferentes sobrecargas de la memoria inmediata. Normalmente, ante un texto más largo se presta más atención al significado y se pasan por alto más errores no contextuales (ortografía).

El habla irrelevante reducía la performance en la detección de errores no contextuales (ortografía) y afectaba más la tarea cuando se veían cinco líneas que cuando se veía una única línea.

Posteriores estudios verificaron que la comprensión se veía afectada también por el contenido significativo del sonido irrelevante. Oswald, Tremblay y Jones [Oswald et al., 2000] plantearon un experimento basado en la comprensión de frases y extracción de las ideas básicas combinado con la memoria inmediata. El diseño experimental partía de varias frases completas, cada una formada por 4 ideas simples. Por ejemplo: “Las hormigas en la cocina comían jalea dulce que estaba en la mesa”. Las ideas simples involucradas son: “Las hormigas estaban en la cocina”, “Las hormigas comían jalea”, “La jalea era dulce” y “La jalea estaba en la mesa”. Estas ideas podían a su vez recombinar-

se de a 2, 3 y 4. Un ejemplo de combinación de 3 ideas simples es: “Las hormigas en la cocina comían jalea dulce”. Por cada idea completa se generaban varias frases simples, dobles y triples combinando las ideas simples que la constituían. Había, así, un repertorio de frases combinadas.

El experimento se dividía en dos fases: de *adquisición* y de *reconocimiento*. Durante la adquisición cada frase del repertorio se presentaba por escrito en una tarjeta durante 10 s. Luego de mostrar cada tarjeta se mostraba una tarjeta que contenía colores que el sujeto debía nombrar en voz alta. El objetivo era obligarlo a mantener la frase en la memoria. La fase de adquisición se completaba con tarjetas que contenían una pregunta elíptica. Por ejemplo: ¿Quién? ¿Qué hizo? ¿Dónde? La pregunta debía ser respondida por escrito.

La fase de reconocimiento constaba de frases ya vistas, nuevas, y anómalas. Las frases *anómalas* eran frases de 4 ideas simples de dos tipos: A y B. Las A combinaban las ideas simples modificando los significados, por ejemplo: “Las hormigas en la mesa comían jalea dulce que estaba en la cocina”. Las B combinaban ideas simples provenientes de diferentes ideas completas. Se mostraban al sujeto las frases y éste tenía que juzgar si la había visto antes, valorando además la seguridad en la respuesta en una escala de 1 a 5.

Los experimentos se hicieron en tres condiciones diferentes durante la adquisición: 1) En *silencio* (grupo de control), 2) Con habla irrelevante *inteligible* y 3) Con habla irrelevante *ininteligible*. El habla inteligible era una grabación de las noticias de la mañana y el habla ininteligible era la misma grabación reproducida retrógradamente. En ambos casos el nivel era de entre 60 dBA y 70 dBA

Se obtuvo que tanto el habla inteligible como el habla ininteligible perturbaban la fase de adquisición. El tipo de errores predominante dependía de la inteligibilidad o no del habla irrelevante. Los errores se dividieron en cuatro clases: 1) Pregunta malinterpretada, 2) Adivinación en lugar comprensión, 3) Igual sentido pero palabras cambiadas y 4) Incapacidad para dar una respuesta. Los sujetos sometidos a habla inteligible cometieron más errores de pregunta malinterpretada. En cambio no hubo efecto estadísticamente significativo para la etapa de reconocimiento.

Los anteriores ejemplos muestran resultados poco concluyentes en relación con estas tareas sencillas y altamente tipificadas. Es probable que en tareas más complejas se pudieran detectar efectos más notorios. Inclusive deberían tenerse en cuenta en la evaluación de los sujetos los resultados de tests de sensibilidad al ruido como el de Weinstein [Weinstein, 1980], ya que en efecto se notan diferencias importantes en la perturbación que el ruido produce en diferentes individuos, sobre todo en tareas intelectuales.

5 Hacia una medida del potencial distractor del ruido

Aun cuando un determinado ruido no conlleve significado, el tener ciertos rasgos comparables a los de sonidos con contenido semántico (por ejemplo la palabra hablada) podría activar mecanismos cognitivos o intelectuales tendientes a intentar decodificar o interpretar esos sonidos (en busca de un presunto mensaje o significado), lo cual desviaría la atención de la actividad intelectual que se estaba realizando. Esto provocaría una disminución del rendimiento en los procesos intelectuales específicos de esa actividad.

El oído recibe información en dos dominios: *espectral* y *temporal*. La información espectral se organiza básicamente en una serie de *bandas críticas* [Zwicker et al, 1999], determinando *una envolvente espectral*, y mediante una extracción de patrón periódico cuando éste está presente (aun en presencia de ruido en el sentido físico), conduciendo a la sensación de tonalidad o evocación de *altura*, por un lado, y a un patrón de resonancias o formantes que lleva a la percepción de cualidades tímbricas.

La información temporal está contenida no en la forma de onda *microscópica* (demasiado rápida para intervenir en procesos corticales) sino en las envolventes energéticas, particularmente las que se encuentran en el rango de 0,1 Hz a 20 Hz. Envolventes más lentas que 0,1 Hz generalmente son difíciles de detectar sin un esfuerzo particular de atención y sobre todo de la memoria, y envolventes más rápidas que 20 Hz ya pueden provocar otras sensaciones diferentes, como la de rugosidad, etc.

Partiendo de este modelo podemos postular que las envolventes espectrales y energéticas podrían ser los rasgos cuya similitud entre el ruido a estudiar y un estímulo con contenido semántico dispara mecanismos cognitivos de semiosis. A partir de la cuantificación de dichas envolventes es verosímil la posibilidad de construir una especie de “índice de distracción” (uni- o multidimensional) que caracterice el potencial distractor de un ruido.

Algo similar se ha hecho en los métodos tipo STI (speech transmission index, índice de transmisión de la palabra) introducidos para la evaluación de la inteligibilidad de la palabra en ambientes ruidosos y/o reverberantes, particularmente el RASTI (rapid speech transmission index) [Steeneken, 1980; Houtgast et al., 1985; IEC 60268-16:1998]. En estos métodos se analiza la reducción de la amplitud de la envolvente en cada banda como indicador de pérdida de inteligibilidad, atendiendo al hecho de que buena parte de la información se encuentra en esa envolvente. (Si el ruido o la reverberación tienden a uniformar las variaciones de amplitud, la inteligibilidad se resiente.)

La idea es, entonces, obtener para cada banda crítica la *envolvente energética*, definida como la raíz cuadrada de la salida de un filtro pasabajos con constante de tiempo de $\tau = 10$ ms ($f_{\text{corte}} = 1/(2\pi\tau) = 15,9$ Hz) en cuya entrada se aplica la señal elevada al cuadrado. Se ha desarrollado un algoritmo que implementa un banco de filtros IIR multitasas utilizando aproximaciones de Butterworth de cuarto orden y descompone la señal en bandas críticas. A modo de verificación, se evaluó la superposición de las salidas de los diversos filtros obteniéndose una recomposición casi indistinguible auditivamente del original. Luego se utilizó un detector de envolvente con constante de tiempo 10 ms, elegida porque permite el paso de la mayoría de las transiciones detectables a la audición. En la figura 4 se muestran dos bandas críticas extraídas de un ruido blanco.

Siempre que se aplica a un ruido blanco un filtro de banda angosta aparece una modulación inherente cuya frecuencia es del orden de $0,64 \Delta f$ (ver [Zwicker et al., 1999]). Dado que las bandas críticas de baja frecuencia son más angostas, dicha modulación inherente es de baja frecuencia, pudiendo ser detectada por el oído. Las bandas más altas son más anchas y por lo tanto la modulación inherente es de mayor frecuencia, de allí la uniformización de la envolvente (el filtro pasabajos del detector las atenúa). El oído, por otra parte, deja de percibir esta modulación debido al fenómeno de post-enmascaramiento.

En la figura 5 se repiten las mismas bandas para una frase emitida por un hablante individual. Vemos un comportamiento completamente diferente al que se tenía para el ruido blanco. Por empezar la envolvente presenta un patrón más definido, con una modulación de menor frecuencia y mayor profundidad de modulación, impresa por la articulación vocal. Por otra parte, la envolvente contiene frecuencias cercanas a los 4 Hz, frecuencia para la cual se percibe la mayor *fuerza de fluctuación* (fluctuation strength) [Zwicker et al., 1999].

Si la forma de onda o el espectro de la envolvente se parece (según algún criterio a determinar) a la forma de onda o al espectro de la envolvente de la misma banda de una señal con contenido semántico de alguna categoría específica (por ejemplo la palabra hablada, la música, el ruido de una máquina, etc.), entonces el cerebro se “sintonizaría” intentando “decodificar” el ruido percibido como si fuera de esa categoría.

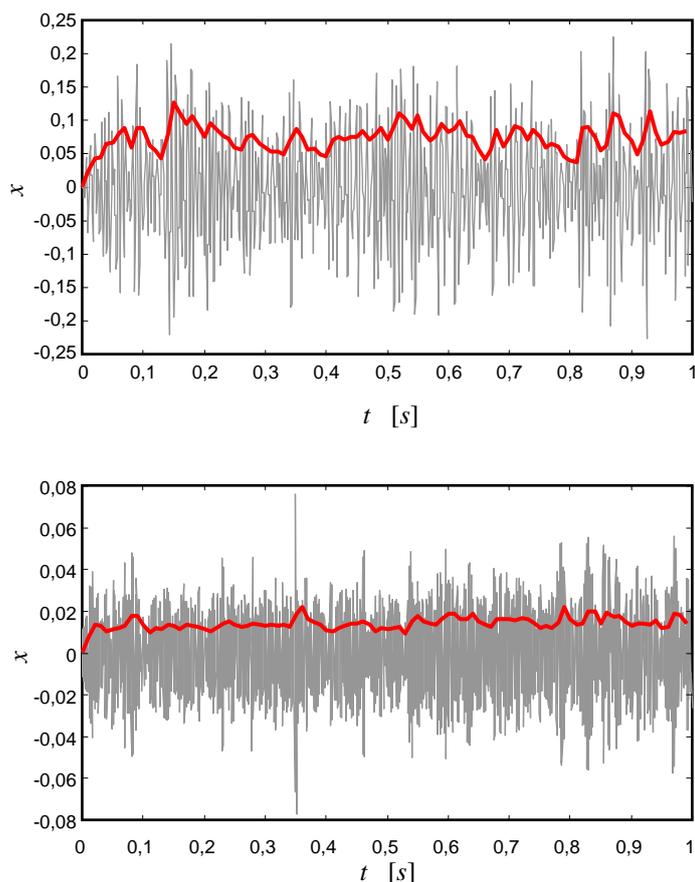


Figura 4. Ruido blanco. Arriba, banda crítica 2 (100 Hz a 200 Hz). Abajo, banda crítica 15 (2300 Hz a 2700 Hz). En rojo, las correspondientes envolventes energéticas.

A efectos de obtener un criterio de similitud se han hecho diversos intentos. El primero, basado en la correlación entre las envolventes de bandas homólogas, resultó poco satisfactorio, obteniéndose correlaciones poco significativas entre señales de una misma clase, por ejemplo, del habla. Ello se debe a que la forma de onda de las envolventes depende bastante de la cadencia y expresividad del hablante. Una alternativa podría ser una alineación temporal previa a la aplicación de la correlación. Esta técnica no fue ensayada.

Probablemente la evolución lenta de las envolventes energéticas implique que su forma de onda sea más importante que su espectro. En otras palabras, al no haber un sistema mecánico como la cóclea (o de otra naturaleza) para descomponer espectralmente estas envolventes, la corteza cerebral se ocuparía de la forma de su evolución temporal más que de su espectro. Sin embargo, el contenido espectral caracteriza ciertos rasgos de la evolución temporal. Por ejemplo, una envolvente de corta duración como en las consonantes oclusivas, tiene contenido de “alta” frecuencia (siempre teniendo en cuenta que “alta” en este caso es en términos relativos al rango de 0,1 Hz a 20 Hz).

Por esta razón se intentó utilizar el espectro y el cepstro. Se compararon señales de voz entre sí (señales similares) y de voz con ruido blanco (señales diferentes). La comparación de los cepstros dio en todos los casos correlaciones muy altas con diferencias poco significativas entre señales similares o diferentes. En cambio, la correlación entre los espectros reales (obtenidos por transformada rápida de Fourier) de las dos en-

volventes arrojó resultados significativamente más altos entre señales similares que entre señales diferentes. En la tabla 1 se muestran dos ejemplos; uno de ellos compara dos alocuciones de 10 s, una realizada en condiciones prácticamente anecoicas, la otra con reverberación; el otro, compara una alocución con ruido blanco. Aunque la promediación directa de las correlaciones de las diversas bandas no necesariamente sea la mejor medida global (ya que podría haber bandas con menor peso en la determinación de la similitud), se proporcionan los valores para ambos casos junto con el desvío estándar. Como se ve, la variabilidad es suficientemente baja para que la diferencia observada en las medias sea significativa.

Hasta el momento no se pudo dar una explicación completa del porqué de las correlaciones relativamente altas entre el ruido blanco y la voz. Posibles motivos pueden ser la modulación inherente y otros efectos del filtrado que producen ondulaciones en el espectro que podrían ser comunes a todas las señales filtradas.

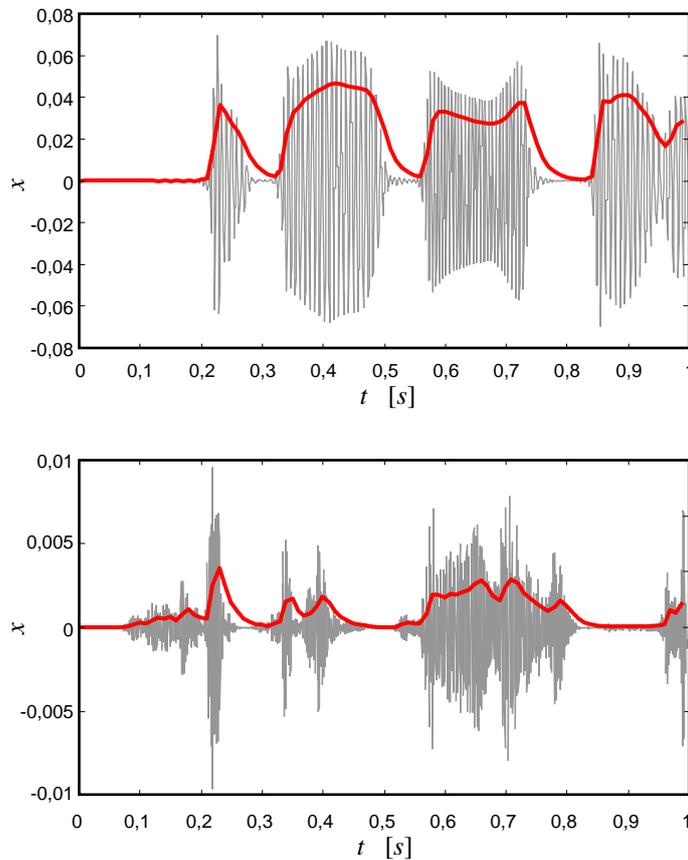


Figura 5. Frase “Se plantea y funda...” pronunciada por un hablante individual. Arriba, banda crítica 2 (100 Hz a 200 Hz). Abajo, banda crítica 15 (2300 Hz a 2700 Hz). En rojo, las envolventes energéticas.

6 Conclusiones

En el presente trabajo se ha fundamentado la necesidad de tener en cuenta el contenido semántico del ruido al haber suficientes indicios de que el mismo afecta la manera en que el ruido afecta a las personas, sobre todo en lo que hace al desempeño en actividades cognitivas. Luego se ha propuesto y experimentado un enfoque para extraer

características comunes a ruidos perceptivamente similares basado en las envolventes espectrales de la señal filtrada con filtros de ancho de banda crítico y se han comparado bajo este punto de vista señales diferentes y similares. El enfoque se muestra promotor, aunque todavía es necesario profundizar el análisis hasta llegar a algún repertorio de señales estándar contra las que comparar una señal genérica, así como una función de valoración que permita correlacionar los efectos con la determinación objetiva.

Tabla 1. Correlaciones entre los espectros de las envolventes

Banda crítica	$\rho_{\text{voz-voz}}$	$\rho_{\text{voz-ruido}}$
1	0,75241	0,64951
2	0,80552	0,68296
3	0,75093	0,65554
4	0,79366	0,62399
5	0,76398	0,73354
6	0,76949	0,68003
7	0,84884	0,71049
8	0,79613	0,68856
9	0,78554	0,68218
10	0,78743	0,74642
11	0,82365	0,69143
12	0,84995	0,68822
13	0,78935	0,67939
14	0,79271	0,69912
15	0,80217	0,67842
16	0,79067	0,66717
17	0,79909	0,65261
18	0,79410	0,6795
19	0,77500	0,68489
20	0,81207	0,65979
Valor medio	0,79413	0,68169
Desvío estándar	0,02633	0,02793

NOTA: El presente trabajo está basado en [Cabanellas et al., 2006]

Referencias

- Cabanellas Susana; Pasch, Vivian; Yanitelli, Marta; Miyara, Federico. "Estudio exploratorio de algunos parámetros perceptivos del ruido vinculados al contenido semántico". FIA 2006, Santiago, Chile, 25/10/06-28/10/06. Actas V Congreso Iberoamericano de Acústica (comp. por Jorge Arenas).
- Colle, H.A., Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behaviour* 15. pp 17-32
- Fuentes, J. L. (1988). Gramática moderna de la lengua española. Bibliográfica Internacional. Madrid.
- Houtgast, T., Steeneken, H.J.M. (1985). A review of the MTF concept in room acoustics and its use for estimating Speech intelligibility in auditoria. *JASA* 77 (3) pp. 1069-1077.
- IEC 60268-16:1998 Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index.
- Jones, D.M. (1995). The fate of the Unattended Stimulus: Irrelevant Speech and Cognition. *Applied Cognitive Psychology*, Vol. 9 pp. S23-S38
- Jones, Dylan M., Macken, William J. (1995). Phonological Similarity in the Irrelevant Speech Effect: Within- or Between-Stream Similarity?. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 21 No. 1, pp. 103-115.
- Jones, D.M., Miles, C., Page, J. (1990). Disruption of proofreading by irrelevant speech: effects of attention, arousal or memory? *Applied Cognitive Psychology*, Vol. 4, pp. 89-108
- Jones, D.M., Smith, A.P., editors (1992). *Handbook of Human Performance*, vol. 1. Academic Press. London.
- Knudsen, Vern O.; Harris, Cyril M. (1978) "Acoustical Designing in Architecture". American Institute of Physics. USA, 1978.
- Marty, C., Marty, R. (1995). La semiótica. 99 respuestas. Edicial. Buenos Aires.
- Miedema, H.M.E., Vos, H. (1998) "Exposure-response relations for transportation noise". *Journal of the Acoustical Society of America* 104 (6), December. 1998.
- Oswald, C., Tremblay, S. Jones, D.M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory*, 8 (5), pp. 345-350
- Steeneken, H.J.M., Houtgast, T. (1980). A Physical Method for Measuring Speech-transmission Quality. *JASA* 67(1) pp. 318-326
- Weinstein, N.D. (1980). Individual differences in critical tendencies and noise annoyance. *Journal of Sound and Vibration* 68, 241-248
- Zwicker, E.; Fastl, H. (1999) *Psychoacoustics - Facts and Models*. Springer-Verlag. Berlin Heidelberg.