



INTELIGIBILIDAD DEL HABLA

Parte 1

Autor

Ralph Jones
Meyer Sound

Traducción

Ezequiel García Pinilla
Spanish Translator of Meyer Sound

Compaginación y Gráfica
Prof. Francisco Rivadera

Supervisión
Prof. Mario de Oyarbide

Año 2009

Artículos Sobre Inteligibilidad del Habla

Sección 1

Introducción

Muchas personas habrán experimentado la siguiente situación:

Usted va conduciendo su coche, las ventanillas bajas y el sintonizador de radio sonando. Escucha una canción nueva, una que nunca antes ha escuchado y que es interpretada por un artista que también le resulta desconocido. Usted quiere saber el nombre del artista para conseguir el disco. El tema finaliza, el locutor comienza a hablar y...

...usted no puede entenderle debido al ruido de la calle.

Como este simple ejemplo lo ilustra, existe una importante diferencia entre la música y la palabra. El cerebro es capaz de "completar" una buena porción de la información perdida cuando escuchamos música, pues hay un alto grado de redundancia (e.g. si usted no reconoce la línea melódica del bajo en los primeros cuatro compases, seguramente lo hará cuando se repitan.) En cambio, el lenguaje hablado es muy rico en información continuamente cambiante y existe menos redundancia que en la música. Basta con que un pequeño porcentaje de la información se pierda o se distorsione para que al cerebro le resulte difícil descifrar el mensaje.

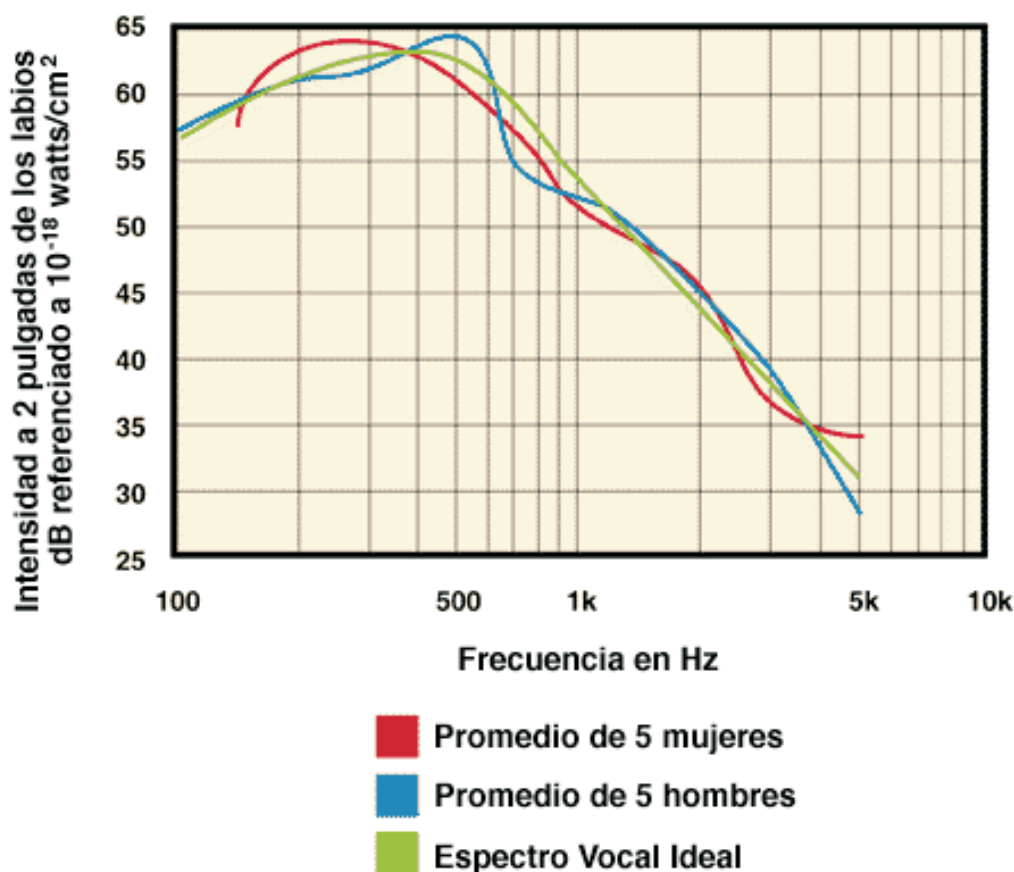
Por lo tanto, los sistemas de audio destinados a la reproducción de palabras están sujetos a requerimientos más rigurosos que aquellos destinados a reproducir música solamente. Estas páginas abordan el tema de la inteligibilidad del habla en los sistemas de refuerzo de sonido. ¿Qué es, qué la afecta y cómo se mide?

La Señal del Habla

El habla humana es una forma de onda continua con una frecuencia fundamental en el rango de 100 a 400 Hz. (El promedio está cerca de los 100 Hz para los hombres y 200 Hz para las mujeres.) A distancias de múltiplos enteros de la frecuencia fundamental, se encuentran una serie de armónicos cambiantes a los que podemos llamar "formantes" y que están determinados por las características de resonancia de la región vocal. Los "formantes" crean los variados sonidos vocales y las transiciones entre ellos. Los sonidos consonantes, que son impulsivos y/o ruidosos, tienen lugar en el rango de los 2Khz a los 9 Khz. En la página siguiente se puede ver un gráfico con el espectro vocal para hombres y mujeres, con un espectro vocal humano "ideal" superpuesto:

La potencia sonora del habla es llevada por las vocales, cuya duración promedio es de 30 a 300 milisegundos. La inteligibilidad es dada por las consonantes, cuya duración promedio es de 10 a 100 milisegundos de duración y pueden ser unos 27 dB más débiles en amplitud que las vocales. La potencia de la señal hablada varía como un todo, y la potencia de los rangos de frecuencia individuales varía con respecto a los otros según cambian los "formantes".

Espectro Vocal Promedio de Oradores Masculinos y Femeninos



Comprensión del Habla

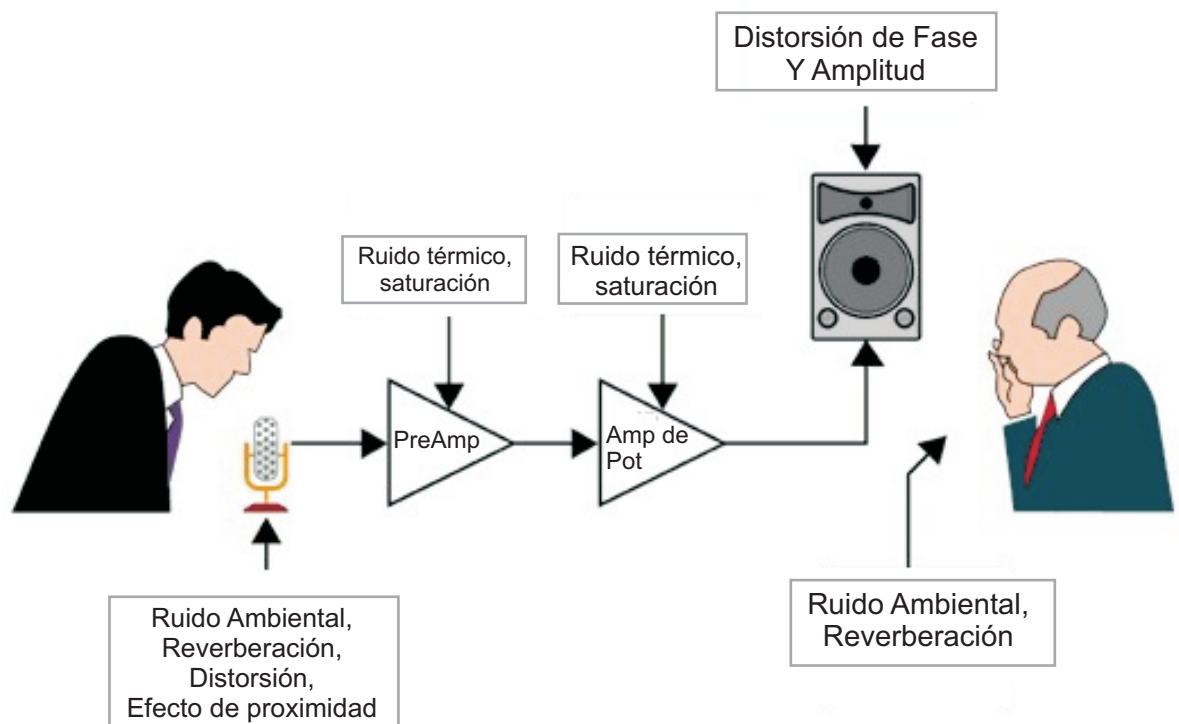
El desafío del oyente es dividir los sonidos del habla en unidades de lenguaje significativas- una tarea complicada, por cierto. Los intervalos en el sonido no necesariamente corresponden a divisiones de palabras o de sílabas. Los sonidos del habla tampoco son eventos discontinuos: generalmente se combinan y se superponen en el tiempo, y la articulación de un fonema dado difiere según el contexto y el orador.

En realidad, la manera precisa en que el mecanismo oído-cerebro decodifica el habla continúa siendo en parte un misterio. Los factores como el ruido, la duración y el contenido espectral ciertamente afectan la percepción del habla, pero su interacción aún no está del todo comprendida. La disminución de la inteligibilidad está asociada con una pérdida de información que se ha codificado en una cantidad bastante alta de elementos interactuantes, y muchos factores influyen en ella. Los ruidos de fondo pueden enmascarar el habla. Tanto la dirección desde donde el oyente percibe la fuente, y la dirección del ruido que la interfiere pueden alterar el grado del enmascaramiento. La inteligibilidad también se ve afectada por lo predecible del mensaje, por la articulación del orador y, no menos, por la agudeza auditiva del oyente.

Sección 2

Factores que Afectan la Inteligibilidad en los Sistemas de Sonido

El objetivo de un sistema de refuerzo de sonido destinado a reproducir la voz humana es entregar ésta a los oyentes con suficiente claridad para ser comprendida. Dada la complejidad de la señal de la voz humana, la tarea de proveer una alta calidad al amplificarla en el mundo real y especialmente en condiciones menos que ideales, es doblemente complicada.



Aquí observamos el diagrama de un sistema de refuerzo de sonido. El mismo nos muestra los principales factores que afectan la inteligibilidad. Nos indica, además, un número de factores acústicos, electromecánicos y electrónicos que debemos considerar para mantenerla. Para poder ocuparse de todos estos factores eficazmente, debemos comprender cómo cada uno de ellos afecta la señal del habla.

Enmascaramiento

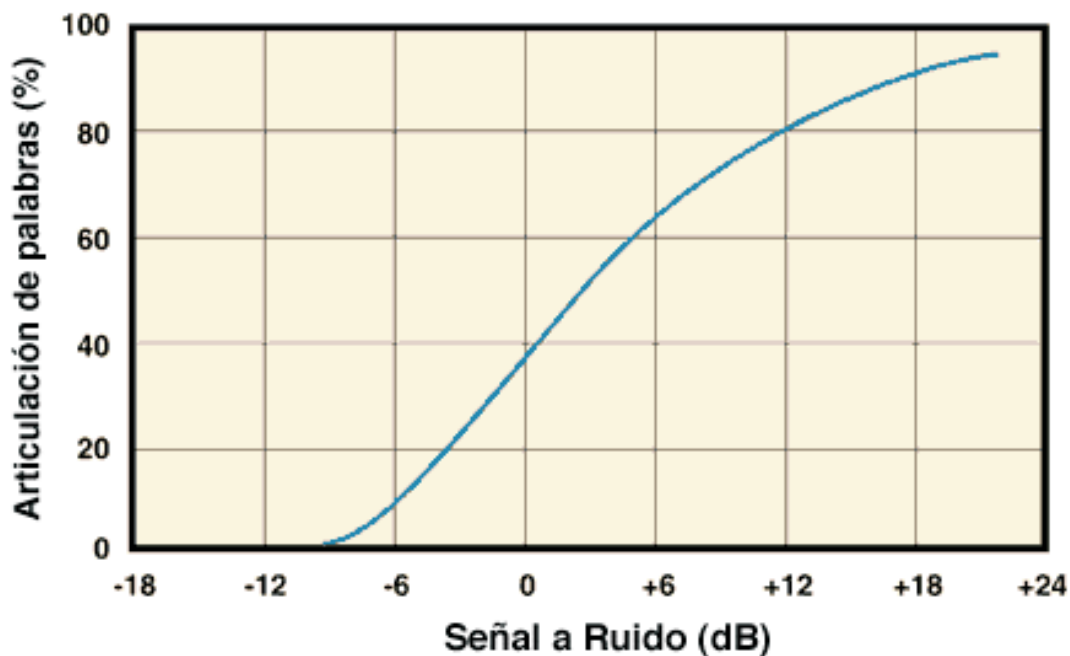
El obstáculo más común que los diseñadores de sistemas de refuerzo de sonido enfrentan es la intrusión de sonidos indeseados que inevitablemente interfieren con la señal del habla. Este efecto es llamado "enmascaramiento", -un término general que comprende una gran variedad de situaciones.

Aquellos ruidos que enmascaran provienen de fuentes acústicas como equipamiento de ventilación, tráfico, muchedumbres y comúnmente, reverberaciones y ecos. Las fuentes electrónicas pueden ser el ruido térmico, el siseo de cintas magnéticas, o distorsiones. Si el sistema de sonido tiene picos muy grandes en su respuesta de frecuencia, la misma señal hablada puede incluso enmascarse a sí misma.

La proporción entre la potencia de la señal del habla y el sonido que la enmascara es llamada "relación señal a ruido" y se expresa en decibeles. Idealmente, esta relación señal a ruido es mayor a 0 dB, indicando que la señal del habla es más potente que el ruido. Cuánto más potente necesita ser la señal del habla para que pueda ser comprendida depende, entre otros factores, del tipo y del contenido espectral del ruido que la enmascara.

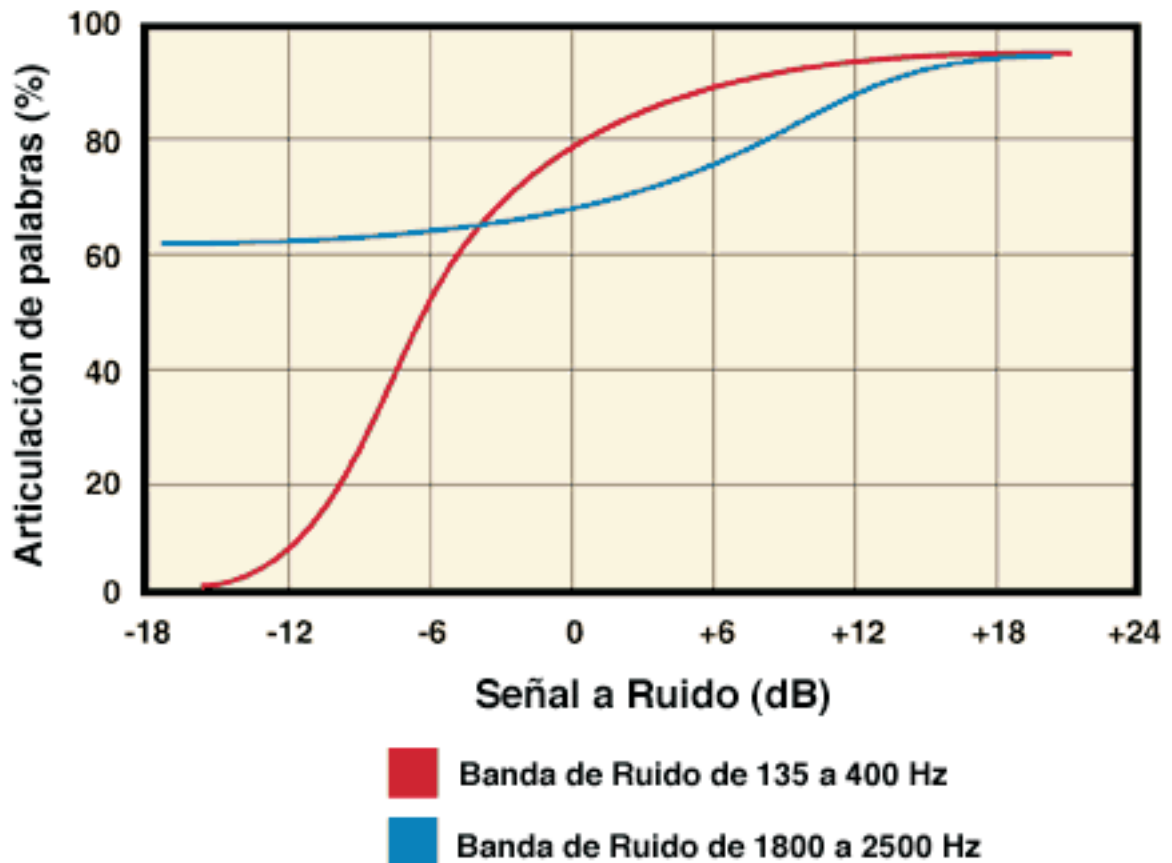
El ruido llamado "de banda ancha" es el que más uniformemente enmascara una señal. Aquí podemos observar un gráfico que nos muestra la articulación de palabras contra la relación señal a ruido cuando la fuente enmascarante abarca frecuencias desde 20 Hz hasta 4 KHz. Nótese que la señal debe ser 12 dB más potente que el ruido de banda ancha para lograr un 80% de reconocimiento de las palabras.

*Articulación de Palabras vs. Señal a Ruido:
Efecto del Ruido de Enmascaramiento de Banda Ancha*



Aunque el ruido de banda estrecha enmascara menos eficazmente el habla que el de banda ancha, el grado de enmascaramiento varía con la frecuencia. Aquí podemos ver una gráfica de la "Articulación de palabras" contra la Relación señal a ruido en dos bandas- 135 a 400 Hz (el rango fundamental de la voz humana) y 1800 a 2500 Hz (el rango en el cual las consonantes son más potentes).

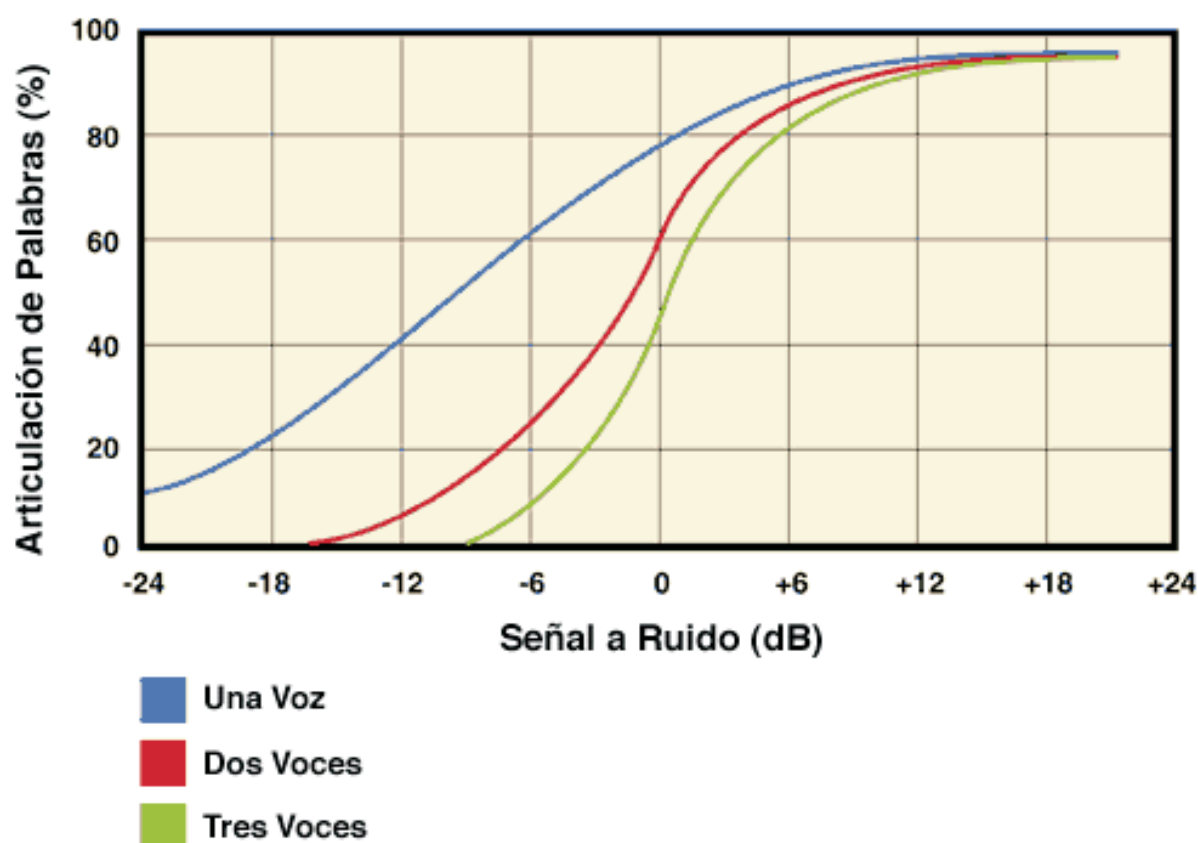
*Articulación de Palabras vs. Señal a Ruido:
Efecto del Ruido de Enmascaramiento de Banda Limitada*



Las altas frecuencias enmascaran sólo las consonantes, y su eficacia como máscara decrece a medida que el ruido se hace más potente. En cambio, los ruidos de bajas frecuencias son más eficaces como máscara cuando el ruido es más potente que el habla, y a altos niveles de presión sonora afectan tanto las vocales como las consonantes. Esta es la causa por la cual el efecto de proximidad de los micrófonos cardioides puede ser tan perjudicial para la inteligibilidad del habla: causa que la señal se enmascare a sí misma. Mientras que los micrófonos cardioides son muy útiles para minimizar la captación de ruido cercana a la fuente sonora, deben ser utilizados con un filtro pasa altos bastante pronunciado (12 dB/ octava o mayor) centrado alrededor de 100 Hz (o más arriba, si el rango de la voz del orador lo permite) para que el efecto de proximidad sea minimizado.

Una voz humana que emite un mensaje que compite con aquel otro que nos interesa escuchar- podemos llamarla "perturbadora"- es también muy buena para enmascarar el habla- particularmente en niveles iguales o inferiores a 0 dB de relación señal a ruido. Además, el efecto de enmascaramiento aumenta con el número de voces "perturbadoras". El siguiente diagrama compara el enmascaramiento producido por una, dos y tres voces. Nótese que, por debajo de 0 dB de relación señal a ruido, tres voces actúan tan eficazmente enmascarando al igual que un ruido de banda ancha. Por encima de 0 dB, sin embargo, la inteligibilidad mejora rápidamente según aumenta la relación señal a ruido. Esto nos ilustra acerca de la importancia de disponer de suficiente potencia para que el sistema de sonido supere el ruido de una multitud.

*Articulación de Palabras vs. Señal a Ruido:
Efecto de Voces Distractoras en Competencia*



La dirección desde la cual el sonido enmascarante llega, relativa a la dirección de la señal hablada, puede afectar el grado de enmascaramiento. Si el ruido proviene del mismo lugar, el enmascaramiento es mayor; disminuye cuando la distancia entre la fuente de ruido y la señal hablada aumenta pues esto facilita al cerebro la discriminación entre ambas. El enmascaramiento es más bajo en el caso de que la exposición sea escuchada con auriculares, el habla en un oído y el ruido en el otro. (Desafortunadamente, no podemos aplicar este mecanismo en los sistemas de refuerzo de sonido).

De lo dicho hasta ahora, podemos concluir porqué la reverberación es tan destructiva para la inteligibilidad, especialmente más allá de la distancia crítica. Siendo causada por el habla, la reverberación imita el espectro de ésta, pero generalmente con mayor energía en bajas frecuencias. Una porción suficiente de reverberación y eco - tal como encontramos en catedrales y grandes estadios deportivos - pueden realmente funcionar como múltiples voces "perturbadoras". Además, por su naturaleza, la energía reverberante llega desde todos los ángulos y, por lo tanto, es difícil de separar de la señal hablada utilizando indicadores direccionales.

Respuesta de Frecuencia

Uno de los factores más obvios en el funcionamiento de los sistemas de sonido que afecta la inteligibilidad es la respuesta de frecuencia. Aquellos sistemas cuyo ancho de banda es extremadamente limitado reproducen el habla pobremente. Por ejemplo, los teléfonos están por lo general limitados a un ancho de banda de 2 KHz, y ello dificulta la distinción entre el sonido de las "f" y la "s" o las "d" y las "t".

Los sistemas de alta fidelidad para reproducción de voces necesitan abarcar el rango desde aproximadamente 80 Hz (especialmente para las voces masculinas profundas) hasta aproximadamente 10 KHz (para una buena reproducción de consonantes, que son cruciales para la inteligibilidad). La respuesta por debajo de 80 hz debe ser atenuada lo máximo posible: no sólo estas frecuencias se encuentran fuera del rango de la señal del habla, sino que además pueden causar especialmente enmascaramiento, que puede ser muy destructivo a altos niveles de presión sonora.

También es muy importante que la respuesta del sistema sea razonablemente plana a lo largo de todo su rango. La atenuación gradual que presentan en altas frecuencias algunos sistemas de sonido en los cuales se prioriza la reproducción de música grabada, puede causar una disminución de la presencia de las consonantes que de por sí, son aproximadamente 27 dB menos potentes que las vocales. Asimismo, picos o depresiones muy pronunciadas en la respuesta de frecuencia del sistema pueden causar tanto auto-enmascaramiento como pérdida de articulación de las consonantes.

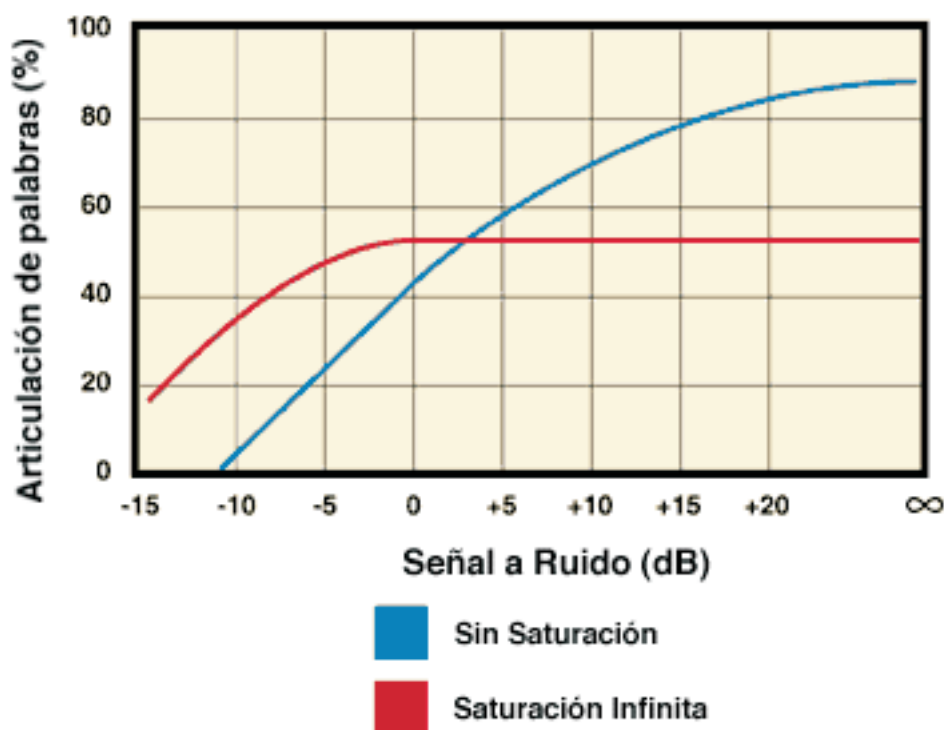
Finalmente, la cobertura del sistema debe ser consistente a lo largo del área que ocupan los oyentes, con cancelaciones mínimas en la respuesta o atenuación fuera del eje en las frecuencias altas críticas. Estos requerimientos muchas veces dictaminan el uso de sistemas distribuidos o dirigidos muy cuidadosamente y también de sistemas retardados. El uso de altavoces con ángulos de cobertura estrechos, ayudará a elevar la relación señal a ruido entre el habla y los niveles de reverberación.

Distorsión

Estudios recientes sobre la inteligibilidad en sistemas de comunicación sugieren que recortando los picos de la señal y luego amplificándola para recuperar su amplitud pico a pico mejora la inteligibilidad. Este "truco" funciona en ambientes ruidosos pues el recorte genera tonos parciales que están armónicamente relacionados con la frecuencia fundamental (distorsión armónica por saturación) y por lo tanto son menos probables de enmascarar el habla- y porque acentúan las consonantes e incrementan la potencia sonora de la señal. Ello ha sido de ayuda para los sistemas de comunicación de banda limitada que son utilizados en ambientes muy ruidosos, como por ejemplo, la cubierta de un portaaviones.

El hecho es, sin embargo, que saturar la señal para mejorar la inteligibilidad funciona solo en los casos donde la relación señal a ruido es muy pobre. Aquí podemos ver un gráfico que nos muestra la articulación de las palabras contra la relación señal a ruido para una señal saturada infinitamente y para una señal no saturada. Nótese que el valor de inteligibilidad para la señal saturada se acerca a un 50 % a 0 dB de relación señal a ruido; cerca de los +3 dB de relación señal a ruido, la señal no saturada logra un valor mejor.

*Articulación de Palabras vs. Señal a Ruido:
El Efecto de la Saturación*



En el funcionamiento real de sistemas de refuerzo de sonido para palabras, la saturación debe ser evitada. Obviamente, sonará objetable en un sistema de alta calidad. Además incrementará el enmascaramiento producido por cualquier sonido que sea captado por el micrófono, ya que ese ruido será saturado junto con la señal del habla.

Otro tipo de distorsión que es muy destructiva para la inteligibilidad es la distorsión por intermodulación. Mientras que ésta puede ser controlada fácilmente en la electrónica del sistema, un grado significativo de IM puede ser generado cuando algunos tipos de altavoces (particularmente los coaxiales de 2 vías) son operados a altos niveles de potencia. La IM genera sumas y diferencias de productos que no están armónicamente relacionados a la frecuencia fundamental. De este modo tienen un efecto de máscara mayor que los productos armónicos generados por la saturación.

Respuesta relativa al tiempo

Quizás porque es escasamente comprendida y sus efectos son más sutiles, la respuesta de fase en sistemas de comunicación ha recibido escasa atención. De hecho, muchas de las publicaciones acerca de "fase" e inteligibilidad, en realidad tratan sobre los efectos de la polaridad relativa. Se ha observado, por ejemplo, que cuando la señal del habla es escuchada con ruido en auriculares, la inteligibilidad aumenta alrededor del 25% si se invierte la polaridad de esa señal en un oído, con respecto al otro. Este efecto no tiene aplicación en los sistemas de refuerzo de sonido, excepto en los monitores intra-auriculares.

*Sección 3***Mediciones Estadísticas de la Inteligibilidad del Habla**

Las mediciones estadísticas de la inteligibilidad para evaluar los sistemas de comunicación, son realizadas con personas, más que con instrumental electrónico.

Estos tests se plantearon en 1910 y se perfeccionaron con la introducción del teléfono y el advenimiento de los sistemas electrónicos de comunicación en la segunda guerra mundial. Son considerados como los métodos más precisos y confiables en lo que a mediciones de la inteligibilidad se refiere. Aún cuando algunas variaciones se han implementado, este apartado está más relacionado con el procedimiento aprobado por el American National Standards Institute (ANSI S3.2-1989, "Método para la Medición de la Inteligibilidad de la Palabra en Sistemas de Comunicación").

El método y sus aplicaciones

El proceso del método estadístico utiliza oradores entrenados, con un vocabulario fluido y que pronuncian listas estandarizadas de palabras a través de sistemas de comunicación. Estas son escuchadas por personas cuya capacidad de audición es también fluida. Las palabras que se incluyen en las listas son seleccionadas para evaluar aspectos específicos de la transmisión del lenguaje hablado; la capacidad de los oyentes para identificar palabras en forma individual o por pares indica la calidad de la transmisión.

Estos tests son utilizados en una amplia gama de aplicaciones. Estas pueden ser tanto examinar las propiedades acústicas de una sala de conferencias, como el funcionamiento de intercomunicadores para buzos que trabajan a grandes profundidades. En sistemas de refuerzo de sonido profesionales, las estadísticas proveen información crucial para arquitectos y consultores, tanto para diseñar los sistemas como para perfeccionar su funcionamiento en el terreno. También pueden ser utilizadas para evaluar las contribuciones que realizan a la inteligibilidad determinados micrófonos, altavoces y procesadores de señal.

Preparación

Para que los resultados de un test sean válidos, aquellas personas que lo dirigen deben estar versadas en el diseño experimental y en el análisis de datos estadísticos. Siendo fundamental el papel que juegan los seres humanos objeto de las mediciones, los experimentadores deben también comprender los factores psicológicos involucrados, incluyendo los efectos de la motivación y el aprendizaje a través de la repetición. Finalmente deben, por supuesto, saber cómo operar apropiadamente el sistema de sonido para evitar la introducción de errores. Por todas estas razones, los tests de inteligibilidad son llevados a cabo por consultores altamente especializados en la materia.

En los tests se utilizan como mínimo cinco oradores y cinco oyentes; grupos más numerosos reducen el margen de error. Los oradores y los oyentes son seleccionados para asegurar la representatividad del sector de edad y sexo al que pertenecen. Todos deben hablar el idioma con el cual se realizará la evaluación (Inglés en este caso) como su lengua materna y tener una capacidad auditiva normal. Los oradores deben tener una buena articulación, y son entrenados para hablar a un volumen consistente y al mismo tiempo sincronizar sus palabras con señales de tiempo, de modo que la velocidad de su exposición no distorsione los resultados de ninguna manera. Los oyentes deben poder discriminar claramente lo que escuchan, además deben estar familiarizados con las palabras utilizadas en el test, con el sonido de la voz de cada orador y con el método de registro de las respuestas.

Existen listas de palabras especiales, que son habitualmente utilizadas para evaluar diversos aspectos de la comunicación de la voz humana. El estándar ANSI especifica tres:

La Prueba de Diagnóstico por Rima Modificada (MRT):

Es una lista de palabras para pruebas estadísticas de inteligibilidad. La *Prueba de Rima Modificada* utiliza 50 listas de seis palabras de rimas o de palabras Inglesas *monosilábicas homófonas* (ej: “went, hold, sent” etc). Cada palabra está construida a partir de una secuencia de sonidos *consonante-vocal-consonante*, y las seis palabras en cada lista difieren únicamente en el sonido de la consonante inicial o final. A los escuchas se les muestra una lista de seis palabras y después se les pide que identifiquen cual de las seis es hablada por el orador.

La Prueba de Diagnóstico por Rima (DRT):

De manera semejante a la *Prueba por Rima Modificada*, la **DRT** usa palabras Inglesas monosilábicas que son contruidas a partir de una secuencia de sonidos consonante-vocal-consonante. En la DRT, ciento noventa y dos palabras son ordenadas en noventa y seis parejas de rimas que difieren únicamente en sus consonantes iniciales. A los escuchas se les muestra una pareja de palabras, y después se les pide que identifiquen que palabra es presentada por el orador.

El conjunto de las 20 listas de Palabras Fonéticamente Balanceadas (PB):

El conjunto de veinte listas de *palabras fonéticamente balanceadas (PB)* fue desarrollado durante la Segunda Guerra Mundial y han sido ampliamente usadas desde entonces en pruebas estadísticas de inteligibilidad. Las palabras de cada lista son presentadas en un orden aleatorio y nuevo cada vez que son habladas dentro de la misma oración portadora.

La prueba PB de inteligibilidad requiere de mayor entrenamiento para los escuchas y oradores que otras pruebas estadísticas, y es particularmente sensible a la relación señal a ruido: un cambio relativamente pequeño en la relación señal a ruido causa un gran cambio en la evaluación de inteligibilidad.

Evaluación

Si es posible, el sistema de sonido debe ser evaluado bajo las condiciones reales de uso. Si existieran posibles fuentes de enmascaramiento como ruido de tráfico o sistemas de ventilación, estos deben estar presentes durante la evaluación y ser documentados en el informe. También es importante que la ganancia del sistema esté ajustada a un nivel de presión sonora realmente representativo. Puede utilizarse material previamente grabado siempre y cuando el equipamiento de grabación o de reproducción no introduzca ruido o distorsión significativos.

Como mínimo, se entregarán a cada orador tres listas *PB* o *MRT* - o la lista completa *DRT* - para ser leídas. Cuando el sistema de sonido a evaluar sea solamente uno, las personas entrenadas intervinientes en el test serán primero evaluadas "cara a cara" o en condiciones supuestamente "ideales" de modo de establecer una medición de referencia. Bajo estas condiciones la inteligibilidad debería ser casi perfecta. Estos valores serán utilizados como referencia para comparar con los obtenidos posteriormente en la evaluación del sistema de sonido. Durante la evaluación deberá incluirse la información adicional como por ejemplo la rapidez o precisión de la respuesta de los oyentes y sus opiniones estadísticas acerca del sistema de sonido.

Análisis de los Resultados

Existen varias maneras de analizar la información obtenida. Ello depende de las características particulares de las listas de palabras utilizadas y de las variables evaluadas. Como mínimo, se establecerán valores porcentuales que muestren las ocasiones en que las palabras fueron identificadas correctamente por el oyente. El promedio de estos valores arrojará un nuevo valor único. Si solo son utilizados el *DRT* o el *MRT*, los resultados son ajustados matemáticamente para justificar una presunción (no así en el caso de utilizar el test *PB*). Análisis estadísticos más profundos pueden proporcionar información más detallada acerca del sistema de sonido.

Autor

Ralph Jones
Meyer Sound

Traducción

Ezequiel García Pinilla
Spanish Translator of Meyer Sound

Compaginación y Gráfica
Francisco Rivadera

Supervisión

Prof. Mario de Oyarbide

Año 2009