41º CONGRESO NACIONAL DE ACÚSTICA
6º CONGRESO IBÉRICO DE ACÚSTICA

DESARROLLO DE UN MÉTODO PARA LA OBTENCIÓN DEL NIVEL DE INTELIGIBILIDAD DE UN SISTEMA SONORO USANDO UNA SEÑAL DE VOZ

REFERENCIA PACS: 43.55.+p

Pérez López, Guillermo Evaristo[(1)]; Martín Cruzado, Carlos G.[(2)]; Luna Ramírez, Salvador[(1)].

[(1)]: Dpto. Ingeniería de Comunicaciones. ETSI Telecomunicación. Universidad de Málaga. Campus de Teatinos. 29071. Málaga. España. Tfno: +34 952137186. Correo-e: sluna@ic.uma.es

[(2)]: Genuix. Sistemas Electroacústicos Avanzados. C\ Saint Exupery. 29007. Málaga. España. Tfno: +34 951212351. Correo-e: carlos@genuix.es

ABSTRACT

This paper describes and analyses a specific mathematic method to calculate the quality level of acoustics in a room in real time using only two voice signals: the voice signal sent and the voice signal received. To validate this method it is used acoustic measurements in forty different rooms. To calculate the acoustic levels it has been used the impulse response of forty rooms as IEC 60268-16 norm says. The custom method and the suggested method have been compared. As a result, mainly because of the value of cross-correlation, it is assumed that this method is right.

RESUMEN

En el presente trabajo se muestra y analiza un método matemático para calcular el nivel de inteligibilidad de un sistema sonoro partiendo de la señal de voz emitida y la recibida por la audiencia. Para validar el método propuesto se parte de mediciones acústicas realizadas en 40 salas diferentes. Se han realizado los cálculos de inteligibilidad partiendo de la respuesta al impulso de cada sala, tal como indica la norma IEC 60268-16. Los resultados obtenidos con el método tradicional se han comparados con el método propuesto. La comparación muestra un nivel de correlación que permite validar el método.

1. INTRODUCCIÓN

La propagación del sonido en un recinto hace que las características del mensaje se vean alteradas en muchas de sus características. En aquellos recintos donde el proceso acústico principal es el de emitir y recibir mensajes (y entenderlos), las modificaciones que el recinto ejerza sobre el sonido deben estar muy bien caracterizadas y, sobre todo, limitadas para que el mensaje original pueda ser entendido por el oyente. Es el caso de los denominados recintos de concurrencia pública (aeropuertos, estaciones, pabellones, etc.).

Con este objetivo de limitar la degradación del mensaje en un recinto, existen diversos parámetros acústicos que intentan cuantificar la inteligibilidad de un mensaje, [1][2]. Esencialmente, dichos parámetros miden, de una forma u otra, las diferencias entre la onda original emitida al recinto y la que recibe el oyente.

Algunos parámetros en concreto hacen uso de ciertas características propias de la voz humana. Es el caso de los parámetros de inteligibilidad STI y RASTI, definidos en [4]. Si bien el sonido audible se encuentra entre 20 y 20.000 Hz, la frecuencia en una conversación normal suele estar entre 125 Hz y 8000 Hz, de modo que es en este rango en el que habitualmente se analiza la señal. Además, la señal de voz se puede emular a un sonido modulado en amplitud, con altos índices de modulación, esto es, varía rápidamente entre intervalos de alta energía y otros de escasa o nula energía. Esta frecuencia de modulación varía entre 0.63 y 12.5 Hz.

Habitualmente, las metodologías de medición de los parámetros de inteligibilidad en recintos determinan a partir de una señal estándar conocida (señal sinusoidal, barrido en frecuencia, ruido rosa…), haciendo necesario la interrupción en la emisión de mensajes ocupando y contaminando el recinto acústicamente con las señales de prueba, con las lógicas molestias. Por esta razón, la caracterización del STI ó RASTI en un recinto se hace antes de la apertura del mismo, dando por supuesto que los resultados obtenidos se mantendrán a lo largo del tiempo y con las diversas circunstancias que ocurran.

El presente trabajo presenta un método de medición de parámetros de inteligibilidad donde las señales a analizar son los propios mensajes de interés a emitir por parte del operador del recinto. De esta manera, no sólo no se interrumpe el trabajo habitualmente a desarrollar en la sala, sino que también es posible hacer un seguimiento constante y evolución de la inteligibilidad a lo largo del tiempo.

2. MÉTODO TRADICIONAL, STI y RASTI [4].

Según [4], el cálculo del nivel de inteligibilidad no emplea señales de habla real. El índice de inteligibilidad se calcula siguiendo, a grandes rasgos, 2 pasos:

1. Obtención del MTF (Función de Transferencia de Modulación) en función de la frecuencia. Dicha curva se obtiene a través de mediciones en distintas frecuencias puntuales según

$$m(F)=\frac{m_o}{m_i} \qquad (1)$$

donde *m(F)* es el índice de transferencia de modulación a una determinada frecuencia de modulación F, m_o es el índice de modulación de la onda de salida (es decir, la capturada tras propagarse por la sala), y m_i lo es para la señal de entrada (la señal fuente emitida a la sala). Las medidas de *m(F)* se hacen para múltiples frecuencias de onda portadora y de onda moduladora, tal como se indicará más adelante.

2. Cálculo del índice STI o RASTI a partir de los MTF calculados siguiendo la formulación que se indica en [4].

Una buena inteligibilidad en un recinto se basa en conservar la integridad de las modulaciones de la señal vocal transmitida. Por ello, STI y RASTI están basadas en la medida MTF, ya que la MTF cuantifica el grado de preservación de las modulaciones vocales en las bandas de frecuencia individuales, sabiendo que las reverberaciones y las reflexiones producen una reducción del índice de modulación. Ambos parámetros se calculan con un algoritmo que mide la inteligibilidad con valores que varían de 0 (inteligibilidad nula) a 1 (inteligibilidad perfecta).

La técnica STI es un método basado en las modulaciones de amplitud que ocurren en el habla natural. Para su cálculo, se averigua el MTF usando 7 filtros de bandas de octava (125

Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz) y 14 frecuencias de modulación (0.63 Hz, 0.8 Hz, 1 Hz, 1.25 Hz, 1.6 Hz, 2 Hz, 2.5 Hz, 3.15 Hz, 4 Hz, 5 Hz, 6.3 Hz, 8 Hz, 10 Hz, 12.5 Hz), dando como resultado 98 valores de MTF, a partir de los cuales se obtiene el valor de STI con unos cálculos más elaborados. Por último, resaltar que el método STI asume que el canal de transmisión es completamente lineal. Por esta razón, una medida STI puede ser "engañosa" para ciertas no linealidades del sistema o procesamiento variante en el tiempo.

El método STI era muy costoso desde el punto de vista del procesado, por ello surge el método RASTI. Este método presupone que la mayoría de la inteligibilidad espectral se concentra en ciertas bandas de frecuencia. Por ello, emplea únicamente 9 valores de MTF tras filtrar la señal únicamente por 2 filtros de bandas de octava (500 Hz y 2KHz) y 9 frecuencias de modulación, simplificando así el método STI. Por tanto, el proceso de cálculo del RASTI es el mismo que el del STI pero con sólo 9 valores de MTF.

Ha de mencionarse que para sendos métodos de medida STI y RASTI, los resultados medidos son medias con cierta desviación típica debido a la aleatoriedad del ruido acústico presente en la medición. La desviación típica respecto al valor teórico o ideal depende entre otros factores de la duración de medición elegida, y es conveniente que la desviación típica sea estimada efectuando mediciones repetidas, al menos para un número limitado de condiciones. De este modo, la aplicación del método RASTI está limitada por los factores relativos a la transmisión del habla, el ruido de fondo y la reverberación.

Una vez obtenidos los 9 valores de *m(F)*, el proceso de cálculo del índice RASTI es como sigue. En primera lugar se calcula la denominada relación señal a ruido aparente, definida como

$$\left(\frac{S}{N}\right)_{App} = 10\log_{10}\left(\frac{m(F)}{1-m(F)}\right) \quad \text{[dB]} \qquad (2)$$

donde *m* es el ratio de índices de modulación a las distintas frecuencias. Esta operación equivale a considerar que todas las reducciones del índice de modulación, se deben a un cierto ruido de fondo. Posteriormente se impone que un recorte de los valores obtenidos al rango [-15 +15] dB.

Finalmente, tras realizar un promediado de los 9 valores obtenidos previamente de $(S/N)_{App}$, se obtiene un único valor, normalizado entre 0 y 1, según la ecuación

$$RASTI = \frac{\frac{\sum\left(\frac{S}{N}\right)_{App}}{9} + 15}{30} \qquad (3)$$

3. METODOLOGÍA PROPUESTA

Tomando como referencia los cálculos expuestos en el apartado anterior, basados en [4], se ha desarrollado un método alternativo para la obtención de la Función de Transferencia de la Modulación (MTF), que permita obtener resultados equiparables usando señales de habla en tiempo real. A partir de ese punto, los cálculos posteriores para la averiguación del RASTI son idéntico para sendas metodologías.

El objetivo es desarrollar un sistema que permita evaluar el nivel de inteligibilidad de cada unos de los mensajes emitidos por un sistema sonoro. Para ello, no se usa ninguna señal estándar como excitación de la sala. El sistema tan sólo usa la señal original emitida y la señal captada en la zona de audiencia degradada principalmente por ruido y/o reverberación. De esta

manera no se interfiere en el sistema de sonido sino que actúa de manera paralela a él y en tiempo real. La figura 2 muestra el esquema de trabajo, en el que la sala actuaría como sistema acústico lineal (con una respuesta acústica impulsiva dependiente de las posiciones de la fuente y receptor, *h(t, Tx, Rx)*) que modifica el mensaje original, *x(t)*, para convertirlo en el mensaje que llega al oyente, *y(t)*.

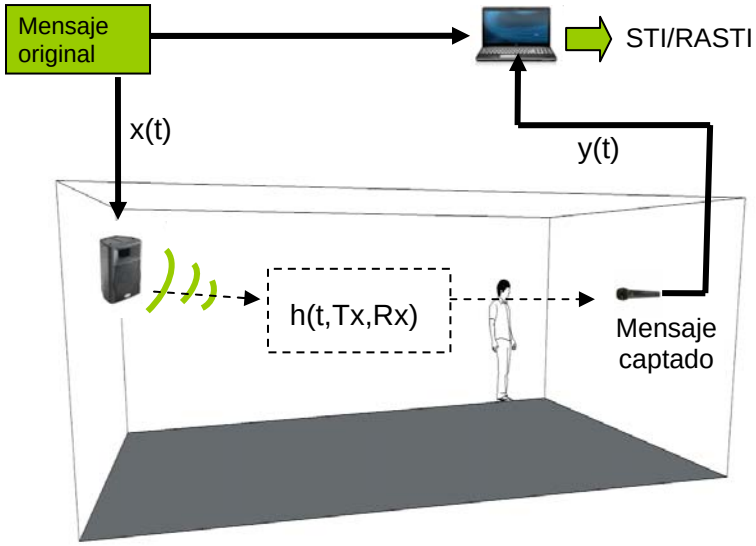

Fig 2.- Esquema de trabajo para el método alternativo

Al igual que en los métodos de medición estándar, en este método básicamente se comparan índices de modulación entre señal emitida y recibida para distintas bandas de frecuencias. La metodología propuesta indefine principalmente un algoritmo para calcular la MTF de una señal de habla de tal modo que permitiera obtener posteriormente el STI o RASTI según los cálculos normalizados a partir del MTF, y en tiempo real. No obstante, se observó la necesidad de añadir un análisis previo de la señal de voz captada. La Figura 3 muestra el esquema de bloques global para el cálculo de la inteligibilidad según la metodología propuesta.

Un primer elemento de este análisis previo en la figura consiste en averiguar la existencia de retardo entre la señal transmitida y la señal recibida, normalmente producido por la geometría de la sala o la electrónica de los sistemas de captación. En su caso, se calcula dicho retardo y se elimina para tener alineadas en el tiempo *x(t)* e *y(t)*. Adicionalmente, es necesario estudiar previamente los índices de modulación de la señal *x(t)* para evaluar si posee suficiente modulación como para permitir los cálculos. De esta manera, no se darán resultados cuando la señal original no tenga suficiente modulación, como pueden ser señales no representativas del habla humana (por ejemplo, un silbido o un ruido constante). En estos casos no procede que el sistema calcule la inteligibilidad del mensaje.

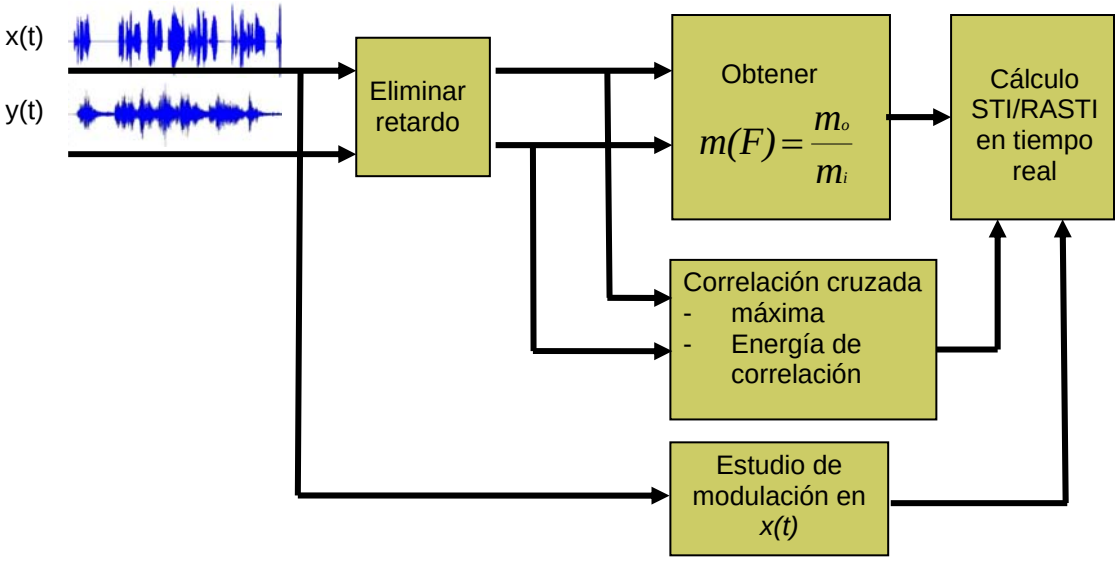

Fig 3. Diagrama de bloques para la metodología propuesta

Un último cálculo consiste en la medida de la correlación cruzada entre *x(t)* e *y(t)*. Este cálculo tiene como objetivo descartar los casos en los que la señal captada en la sala, *y(t)*, no tiene ninguna relación con la original emitida en la sala, *x(t)*. De esta manera, se evitará obtener un resultado de inteligibilidad no asociado al mensaje emitido en aquellos casos en los que, por ejemplo, exista algún transeúnte hablando cerca del micrófono de captación, siendo entonces *y(t)*, obviamente, una señal totalmente ajena al mensaje original emitido. Este proceso, junto con el de estudio de modulación en *x(t)*, descartan dos casos que pueden interferir el estudio de inteligibilidad de la sala, a) una señal original con poca modulación (no representativa del habla) y, b) señal externa captada y ajena a la excitación original.

3.1 Cálculo de *m(F)*

No obstante los cálculos anteriores, el núcleo de la metodología propuesta es el cálculo de los índices de modulación a las distintas frecuencias, *m(F)*. En los métodos STI y RASTI este cálculo es sencillo debido a la naturaleza estándar de las señales de excitación empleadas (senos modulados, generalmente). Cuando la señal de excitación es un mensaje de voz humana, dicho cálculo es distinto, aunque inspirados en el cálculo definido en [4].

El cálculo definido en la normativa toma como parámetros de entrada la respuesta acústica impulsiva del recinto, *h(t)*, y la relación señal a ruido existente. En la metodología propuesta, la señal original, *x(t)*, y degradada, *y(t)*, es la única información necesaria. La figura 4 muestra el proceso de cálculo de los índices de modulación. A partir de las señales *x(t)* e *y(t)*, se analizan de manera paralela ambas señales con, en primer lugar, un filtrado en cada una de las 7 bandas de octava definidas en la norma para el cálculo de STI, y, posteriormente, una detección de envolvente y normalización. Una vez acondicionada la señal, se realiza un nuevo filtrado en 1/3 de octavas para cada una de las 14 frecuencias de modulación definidas en la norma. Finalmente, se obtiene el índice de modulación de la señal original, m_i, y de la señal degrada, m_o, asumiendo como tal índice el valor de la desviación estándar de la envolvente.

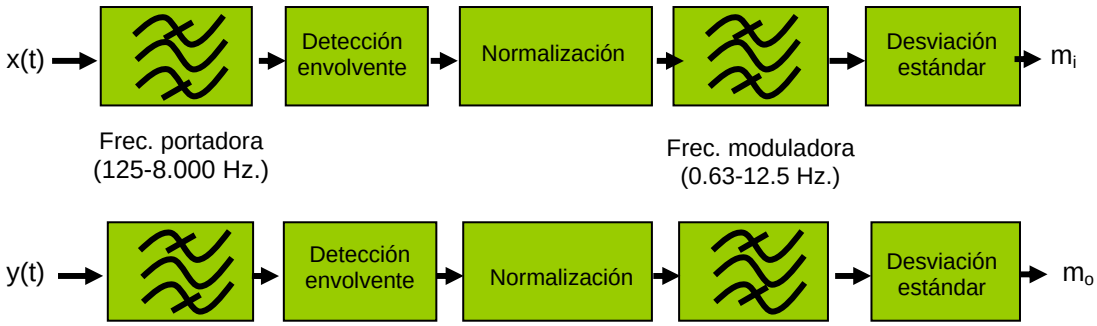

Fig 4. Obtención de los índices de modulación según metodología propuesta

Para terminar el proceso matemático, tan sólo queda obtener la función de Trasferencia de Modulación (MTF) mediante el cociente de m_o/m_i. Este MTF se obtiene para 98 valores, correspondiente a la combinación de las 7 frecuencias sonoras y 14 frecuencias de modulación. A partir de la obtención de los MTF, el cálculo del valor de inteligibilidad es idéntico al expuesto en la norma CEI 60268-16 [4].

4. VALIDEZ DE LOS RESULTADOS

Para validar la metodología propuesta se aplicó el procedimiento definido a 40 pares de medidas *x(t)* e *y(t)*, obtenidas a partir de la aplicación de sendas respuestas al impulso distintas, *h(t),* correspondientes a sendas salas. Dicho procedimiento se repitió para tres excitaciones distintas pronunciando un mensaje sonoro en distintos idiomas (español, inglés y francés), obteniendo así 120 medidas, correspondientes a 3 mensajes en 40 salas distintas. A ese conjunto de medidas se aplicaron ambos procedimientos. El procedimiento estándar necesita de la respuesta al impulso de la sala, *h(t)*, obteniendo así 40 valores de inteligibilidad. El procedimiento propuesto en este trabajo necesita de los pares de señales *x(t)* e *y(t)* (120 valores de inteligibilidad).

La figura 5 muestra la comparación entre el índice RASTI obtenido por el método estándar y obtenida según la metodología propuesta, colocados en el eje horizontal y vertical, respectivamente. Los 3 colores indican los respectivas señales *x(t)* que se usaron. La coincidencia con la diagonal dibujada indicaría valores idénticos para ambos métodos. Para valorar cualitativamente si estas desviaciones son lo suficientemente bajas, se ha realizado en la figura 6 un gráfico similar pero, en este caso, comparando los valores STI y RASTI obtenidos ambos con el método estandarizado en [4] para las 40 salas disponibles. Cabe recordar que el índice RASTI es una simplificación en el cálculo con respecto al método STI, si bien ambos son ampliamente considerados buenos evaluadores del nivel de inteligibilidad.

Para su mejor comparación numérica, la Tabla 1 compara las magnitudes estadísticas de error y correlación en ambas figuras, poniendo de manifiesto que la desviación de la metodología propuesta es similar, o incluso mejor, a la de la metodología estándar. Esto significa que el posible error cometido al calcular la inteligibilidad con el método propuesto con respecto al nivel STI correcto es, al menos, similar al error cometido calculando el índice RASTI con la metodología estándar.

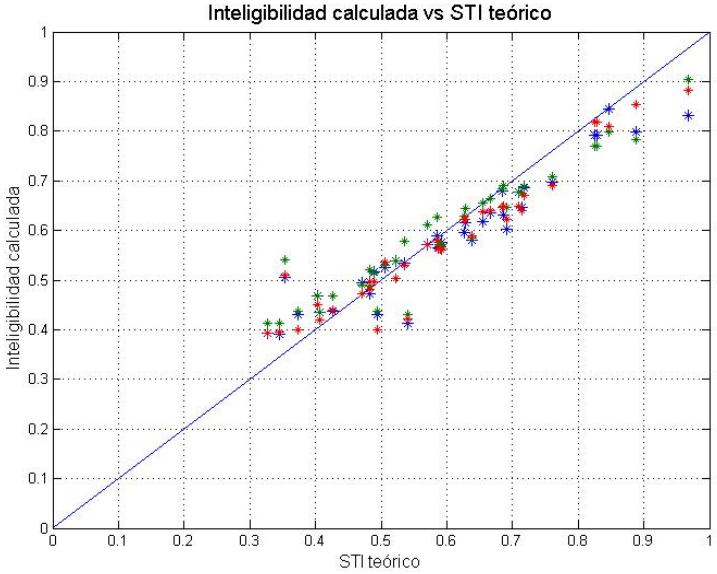

Fig 5. Obtención de los índices de modulación según metodología propuesta

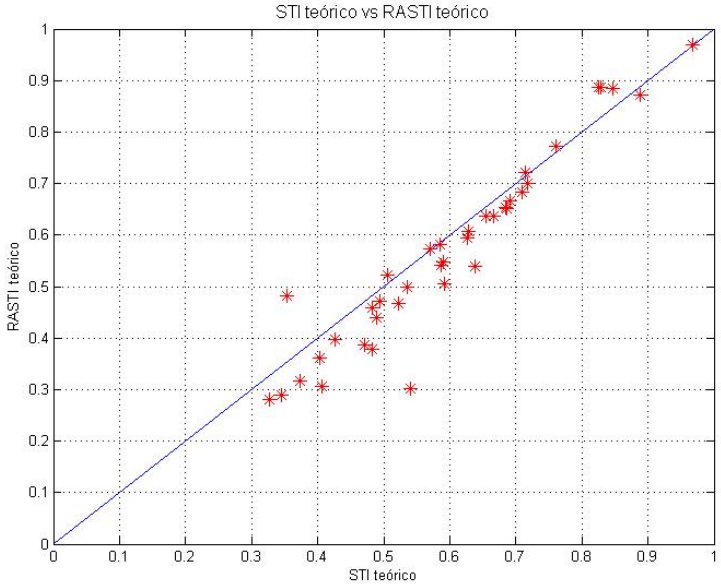

Fig 6. Obtención de los índices de modulación según metodología estándar.

	M. estándar	M. propuesta
Correlación	0.9545	0.9524
Error cuadrático medio	0.0648	0.053
Desviación típica	0.0574	0.0525
Error medio	0.477	0.393

Tabla 1. Comparación de magnitudes estadísticas para ambos métodos.

5. CONCLUSIONES

Se ha presentado un método que permite controlar el grado de inteligibilidad de una sala en tiempo real partiendo de la emisión de una señal de voz humana, frente a los métodos estándar que usan señales patrón como ruido rosa o señales sinuidales. De esta manera, el índice de inteligibilidad podría ser evaluado sin necesidad de interrumpir el servicio del recinto en el que se pretende caracterizar la inteligibilidad. Además, el método permite evaluar en tiempo real los mensajes que, a cada instante, se van emitiendo en la sala. De esta forma, el método propuesto permite evaluar la inteligibilidad a cada instante y para cada mensaje emitido en una sala.

El método propuesto está basado en el estándar definido en [4], aunque variando los cálculos de índice de modulación mediante un proceso de detección de envolvente y cálculo de desviación típica. La validación de dicha metodología se ha hecho cuantificando el error de este método de cálculo sobre los valores teóricos de inteligibilidad en un conjunto de 40 salas y para mensajes en 3 idiomas distintos. Los valores de error obtenidos no superan a los errores obtenidos al aplicar sobre el mismo escenario la metodología estándar, ampliamente integrada en los programas de caracterización acústica de salas.

REFERENCIAS

[1] H. Arau. *ABC de la acústica arquitectónica.* Barcelona: CEAC, 1999.

[2] A. Carrión Isbert. *Diseño acústico de espacios arquitectónicos.* Barcelona: Edicions UPC, 1998.

[3] Ben Gold y Nelson Morgan. *Speech and audio signal processing: processing and perception of speech and music.* John Wiley & Sons.

[4] Norma Internacional CEI 60268-16. *Equipos para sistemas electroacústicos: evaluación objetiva de la inteligibilidad del habla mediante el índice de transmisión del habla.*

[5] Javier Camusso, Cristian Lértora y Federico Miyara. *Obtención del Índice STI a partir de la voz.* VI Congreso Iberoamericano de Acústica – FIA 2008. Escuela de Ingeniería Electrónica, Universidad Nacional de Rosario. Argentina.

[6] J.L. Sánchez Bote, J. González Rodríguez y J. Ortega García. *Array de micrófonos en tiempo real basado en sustracción perceptual y su evaluación mediante E-RASTI.* DIAC E.U.I.T. Telecomunicación. Universidad Politécnica de Madrid. Área de Tratamiento de Voz y Señales (ATVS).