

ANÁLISIS DE INTELIGIBILIDAD Y APROXIMACIÓN A PARÁMETROS DE SALA MEDIANTE INTERNET DE LAS COSAS

Jesus Lopez-Ballester¹, Rafael Fayos-Jordan¹, Jaume Segura-Garcia¹, Santiago Felici-Castell¹, Juan J. Pérez-Solano¹, Maximo Cobos¹, Rosa Cibrián², Alicia Giménez-Pérez³

¹ Dpt Informàtica, ETSE – Universitat de València
Av. Universitat S/N, 46100, Burjassot, València, Spain

e-mail: {jesus.lopez-ballester, rafajor, jaume.segura, santiago.felici, juan.j.perez, maximo.cobos}@uv.es

² Dpt Fisiología, Fac. Medicina – Universitat de València
Avda Blasco Ibáñez, 15, 46010, València, Spain

³ Dpt Física Aplicada, ETSII – Universitat de Politècnica de València
Camí de Vera S/N, 46022, València, Spain

Resumen

En los últimos años, las Redes de Sensores (WASN) se han aplicado ampliamente a diferentes campos acústicos en ambientes interiores y exteriores. La mayoría de estas aplicaciones están orientadas a localizar o identificar fuentes y medir parámetros acústicos específicos del entorno en cuestión.

En este artículo, estudiamos la aplicación de un WASN para las mediciones acústicas de las habitaciones. Para evaluar estas características acústicas se ha utilizado un conjunto de Raspberry Pi 3 (RPi). Se emplea una para reproducir diferentes señales acústicas y cuatro para grabar en diferentes puntos de la sala simultáneamente. Las señales se envían de manera inalámbrica a un equipo conectado a un servidor, donde empleando MATLAB se calculará tanto la respuesta impulsiva (Impulsive Response, IR), como diferentes parámetros acústicos, en particular el Speech Intelligibility Index (SII). De este modo, se han explorado la evaluación de los parámetros acústicos de la sala con mediciones asincrónicas de IR. Por último, se han analizado los requisitos de la red para valorar la eficacia de este sistema.

Palabras clave: WASN, Acústica de salas, respuesta impulsiva, índice de inteligibilidad del habla, estimación de parámetros de sala, estimación de parámetros de habla.

Abstract

In the last years, Sensor Networks (WASN) have been widely applied to different acoustic fields in outdoor and indoor environments. Most of these applications are oriented to locate or identify sources and measure specific features of the environment involved.

In this paper, we study the application of a WASN for room acoustic measurements. In order to evaluate the acoustic characteristics a set of Raspberry Pi 3 (RPi) has been used. One is used to play different acoustic signals and four are used to record at different points in the room simultaneously. The signals are sent wirelessly to a computer connected to a server, where using MATLAB will calculate both the impulse response (IR), and different acoustic parameters, such as the Speech Intelligibility Index (SII). In this way, the evaluation of room acoustic parameters with asynchronous IR measurements two different applications has been explored. Finally, the network features have been evaluated in order to assess the effectiveness of this system.

Keywords: WASN, Room Acoustics, Impulse Response, Speech Intelligibility Index, Room parameters estimation, Speech parameters estimation

PACS n°. 43.55.-n, 43.58.-e, 43.71.-k

1 Introducción

Desde que Sabine [1] publicó sus investigaciones sobre la reverberación en la acústica de salas, muchos investigadores han dedicado muchos esfuerzos a medir el comportamiento acústico en entornos interiores. El procedimiento de medición fue estandarizado en la ISO 3382 [2, 3, 4], y desde entonces se han hecho diferentes revisiones. Esta norma hace algunas recomendaciones sobre diferentes aspectos relacionados con el procedimiento de medición, pero también deja algunos aspectos no obligatorios para permitir la innovación.

Los Redes Inalámbricas de Sensores Acústicos (WASN) han permitido cierta automatización técnica del procedimiento de medición atendiendo a recomendaciones específicas y también siguiendo la norma ISO, pero es necesario seguir investigando para mejorar los procedimientos de medición abordando diferentes técnicas de procesamiento de señal [5, 6, 7, 8, 9] y también los protocolos de medición que utilizan esta tecnología WASN para reunir la información acústica de toda la sala en una sola toma de la fuente de sonido utilizando una serie de sensores distribuidos en la sala.

En los últimos años, las WASN se han aplicado ampliamente a diferentes campos acústicos en ambientes interiores y exteriores. La mayoría de estas aplicaciones están orientadas a localizar una o varias fuentes [10, 11], rastrear [12] o identificarlas [13], y medir características específicas del ambiente involucrado [14]. En [15], los autores utilizan una simulación WASN para estimar el rendimiento de la acústica de la sala, mediante Room Impulse Response (RIRs) localizados y promediados, y adaptan la ecualización multipunto de un sistema de sonido, para renderizar el sonido considerando la inversión del modelo acústico de la sala, y aplicado como pre-filtro a la señal “seca” antes de la reproducción.

La mayoría de parámetros empleados para el análisis acústico de salas, descritos en las fuentes anteriores como Time Reverberation 30 (TR30), TR60, Claridad, podemos sumar el Speech Transmission Index STI y el Speech Intelligibility Index (SII) [16], como parámetros orientados específicamente a la transmisión oral de la palabra. Si bien estos parámetros nos hacen pensar en teatros, cines, conservatorios o salas de producción musical, es muy interesante analizar estos parámetros en salas dedicadas a otros fines, pero donde la transmisión de la palabra resulta vital, como pueden ser los espacios docentes o los dedicados a la atención médica, donde se analizan para evaluar el comportamiento y adecuación de los mismos [17, 18, 19].

Los parámetros mencionados se calculan a partir de la respuesta impulsiva de la sala, no así el SII [20], que se puede extraer de una señal de habla grabada en la posición de la sala en la que deseamos analizar la inteligibilidad. Esto facilita mucho el procesado de señal necesario y por lo tanto la capacidad de cálculo de los dispositivos empleados. Teniendo en cuenta la época actual y los problemas derivados del COVID19 en el ámbito de la educación, con la adaptación de aulas no destinadas a docencia y la distancia social necesaria dentro de las mismas, nos ha parecido de especial interés el hecho de poder disponer de una medida rápida de la inteligibilidad del habla en diferentes puntos de una sala y poder actuar en consecuencia.

Por ello hemos, hemos centrado nuestro estudio en el SII principalmente, además de en el cálculo de los demás parámetros citados. Tal y como hemos mencionado, el cálculo de los mismo es mucho más complejo, por lo que no es posible realizarlo de forma sencilla, rápida ni efectiva en un nodo de nuestra WASN, tal y como sucede por ejemplo con los parámetros de molestia psico-acústica. Por ello nos planteamos implementar una red neuronal convolucional (CNN) que permita predecirlos de manera exitosa, tal y como hicimos también con los parámetros de molestia psico-acústica [21].

Por lo tanto, en este trabajo exponemos el diseño y la implementación de un sistema WASN basado en RPi, la biblioteca PortAudio y MATLAB, para realizar mediciones acústicas síncronas y asíncronas orientadas a estudiar principalmente la inteligibilidad del habla en una sala o parámetro SII, además las reflexiones tempranas en la misma y las características acústicas del espacio.

Si bien el cálculo del SII se puede realizar de forma asíncrona, pues cada nodo arrojará un valor de SII, para el resto de parámetros se han grabado las señales de forma síncrona mediante la conexión WiFi de las RPi. Si bien esta sincronización no es demasiado precisa (debidos a los retrasos introducidos por la red), es más que suficiente para el propósito de este análisis, que nos permitirá tener una idea general del comportamiento acústico de la sala.

2 Materiales y métodos

Para desarrollar las aplicaciones antes mencionadas, hemos utilizado una WASN con 5 nodos formados por la placa Raspberry Pi 3b (RPi), 4 para registrar y una para emitir. Esta plataforma se basa en el sistema Broadcom BCM2837 en chip, que incluye un procesador ARM Cortex-A53 de cuatro núcleos de 1,2 GHz, con 512 KB de caché L2 compartida, una unidad de procesamiento gráfico (GPU) VideoCore IV a una frecuencia de reloj de 400 MHz, con WiFi incorporado (estándar IEEE 802.11 b/g/n) y Bluetooth y 1 GB de RAM LPDDR2 a 900 MHz, con una memoria de tarjeta de micro-SD. Instalamos en los nodos de medición un micrófono de condensador Andoer B01LCIGY8U-USB, con sensibilidad = -30 ± 3 dB, SNR > 36 dB y respuesta de frecuencia lineal entre 50Hz y 16KHz).

La configuración establecida permite que cada nodo esté preparado para adquirir automáticamente un tiempo configurable de audio con 16 bits por muestra (permitiendo un rango dinámico de $20 \cdot \log_{10}(2^{16}) = 96.33$ dB) a una frecuencia de muestreo de 44.1 kHz. Las RPi han sido colocadas en estructuras protectoras, cada una con baterías de 3.7V y 3800mAh. Las medidas de autonomía nos han proporcionado resultados de 10 horas de autonomía aproximadamente, tiempo más que suficiente para realizar el análisis, puesto que éste se realiza en minutos. Es muy útil tener autonomía extra, puesto que en el caso del SII, podremos situar los nodos en distintas posiciones y calcular valores de inteligibilidad en más de 4 posiciones, debido a que no tienen por qué ser simultáneas las medidas.

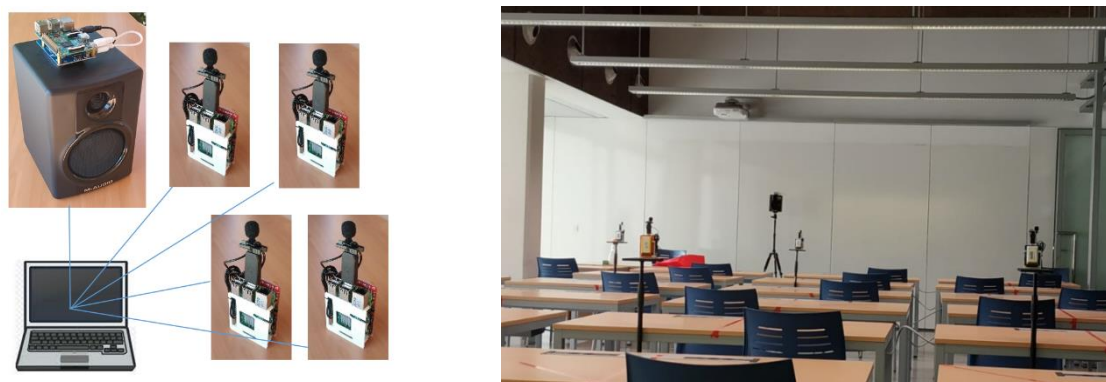


Figura 1 – Diagrama de la arquitectura de IoT para el sistema de medida de parámetros de sala.

La Figura 1 muestra el sistema de medida de parámetros de sala y su empleo en una toma de medidas. Con los nodos de registro y el de emisión configurados, y con un servidor NTP configurado para tener el mismo tiempo en todos los nodos, desde el equipo central, que actúa de servidor, se envía la orden al nodo reproductor para que comience la reproducción y a los nodos grabadores para que comiencen a

grabar al mismo tiempo. En el caso de la reproducción, se puede reproducir tanto un barrido en frecuencia que permitirá la extracción de la IR de la sala y posterior cálculo de diferentes parámetros, como diferentes señales de habla, que permitirán el cálculo directo del SII en la posición de cada nodo. Una vez terminada la grabación, los audios son volcados al equipo central que se encarga de almacenarlos junto con las posiciones de cada nodo y de realizar los cálculos pertinentes.

Para los barridos en frecuencia se utilizan audios de 10 segundos a 1 minuto de duración, pero para el análisis de SII se han obtenido resultados excelentes con audios de 10 segundos e incluso con audios de 5 segundos de habla constante. Cabe mencionar que es interesante hacer un análisis rápido, ya que cuando se apliquen redes neuronales podremos entrenarla y predecir parámetros de manera más eficiente y rápida, por lo que interesa que no sean audios excesivamente largos.

El software ha sido diseñado y probado inicialmente con simulaciones por software de aulas y salas de diferentes geometrías y actualmente el sistema de adquisición y cálculo está operativo, y se ha puesto a prueba en diferentes aulas de la ETSE de la Universitat de València, aprovechando la adaptación de las mismas a la distancia interpersonal requerida para la docencia.

Los parámetros evaluados se muestran en la siguiente Tabla 1. Como se puede ver, en el caso de este conjunto de parámetros, sólo podemos calcular uno (SII) a partir de una señal de habla únicamente. Por ello centraremos la presentación de resultados en este parámetro. El estudio del SII nos puede aportar una idea clara de cómo afecta la posición de la fuente, la distancia, geometría y materiales del aula a la inteligibilidad de la palabra.

Nuestro sistema en conjunto permite medir el SII durante una clase en tiempo real, sin necesidad de interferir en la misma, únicamente grabando al ponente con los nodos situados en las posiciones deseadas.

PARÁMETRO	NECESARIO PARA EL CÁLCULO
RT30 (Reverb. Time 30dB)	Respuesta Impulsiva
RT60 (Reverb. Time 60dB)	Respuesta Impulsiva
ST Early (Componentes ≤ 100 ms)	Respuesta Impulsiva
ST Late (Componentes > 100 ms)	Respuesta Impulsiva
Speech Clarity (C50)	Respuesta Impulsiva
STI (Speech Transmission Index)	Respuesta Impulsiva - IEC 60268-16 2003
RASTI (Rapid Speech Transmission Index)	Respuesta Impulsiva - IEC 60268-16 2003
SII (Speech Intelligibility Index)	Señal de audio, ANSI S3.5-1997

Tabla 1.- Relación de parámetros de sala y cálculos necesarios.

3 Resultados

La mejor sincronización que se ha conseguido con el sistema descrito en la Figura 1 es de 30us, por lo que, pese a que grabemos a 44100Hz nuestro sistema de registro en red sólo permite tener cálculos fiables hasta una frecuencia de la señal de 16 KHz, que es suficiente para el sistema que nos ocupa aplicado a aulas y habla, y que nos ha permitido centrarnos en el SII. De hecho, en SII, por ejemplo, lo interesante es analizar hasta los 8kHz en bandas de 1/3 de octava, que es lo más preciso que describe la norma ANSI S3.5-1997 [20].

3.1 Evaluación de SII

Centrándonos en el SII, se ha desarrollado el algoritmo siguiendo la norma ANSI S3.5-1997. Se muestrea a 44100 Hz y se analizan las 18 bandas seleccionadas de 1/3 de octava del total del espectro audible, como se puede ver en la Figura 2 entre 160 y 7900 Hz. Los resultados mostrados en dicha Figura 2 muestran 2 ejemplos con SII alto y bajo, teniendo en cuenta la escala de inteligibilidad de la Tabla 2.

SII	Inteligibilidad
[0 – 0.3]	Muy mala
[0.3 – 0.45]	Mala
[0.45 – 0.6]	Aceptable
[0.6 – 0.75]	Buena
[0.75 – 1.0]	Excelente

Tabla 2.- Escala de inteligibilidad según parámetro SII.

La Figura 2a corresponde a un registro con SII de 0.79 y una percepción excelente, mientras que la Figura 2b muestra las bandas de un audio con SII de 0.43 y una mala inteligibilidad.

Por lo que podemos ver en esta figura, para el cálculo de SII importan no sólo los niveles SPL por bandas, sino la distribución de los mismos en el espectro, ya que a cada banda se le da un peso específico de importancia según la norma ANSI (las llamadas “One-third-octave frequency importance functions”).

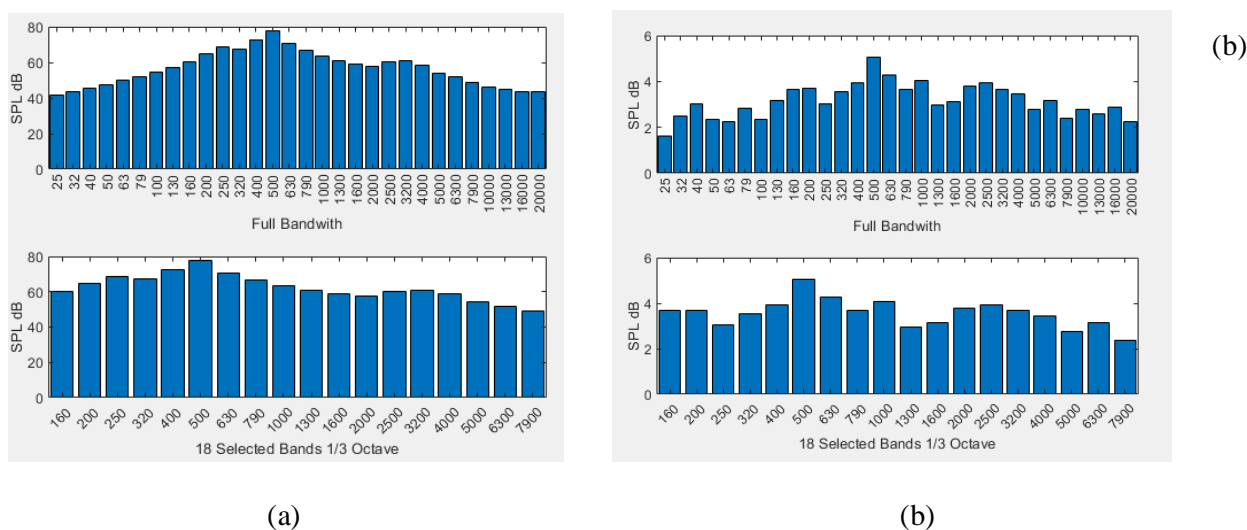


Figura 2 – Espectro de las 18 bandas necesarias para el cálculo de SII.

3.2 Estudio de aplicación del sistema de medida de SII en una sala

Una vez se ha determinado el parámetro de sala a evaluar, se he aplicado al análisis de inteligibilidad en distintas posiciones de un aula para dibujar mapas de SII en función de la posición en que se encuentre el oyente y poder estimar así qué ubicaciones tienen peor percepción del habla.

Se ha desarrollado un sistema de representación que reproduce en 2D y 3D las dimensiones de la sala a tratar, así como los valores calculados de SII para las posiciones indicadas de cada nodo. Se ha preparado el sistema para que sobre-imprima un plano de la sala en el caso de que dispongamos del mismo para facilitar mejor el análisis de los resultados.

En la Figura 3 se muestra un ejemplo del estudio del SII del aula 2.1.1 de la ETSE de la Universitat de València, una sala que tiene unas dimensiones de 8x12x3 metros. En la misma se realiza una representación polar de SII, que nos permite evaluar gráficamente las distancias del hablante y las posiciones a partir de las cuales puede haber problemas de inteligibilidad.

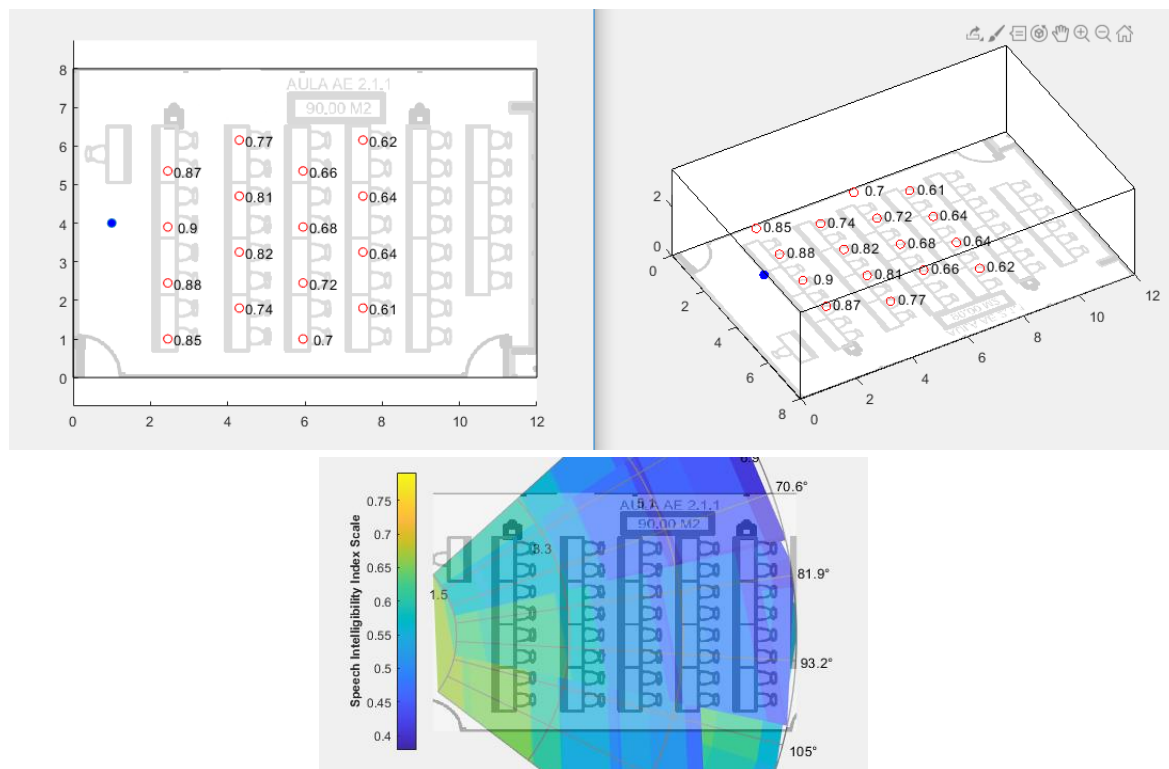


Figura 3 – Mapeo de las medidas de SII en el aula 2.1.1 de la ETSE.

Los resultados mostrados en la Figura 3 nos permiten visualizar como caen los valores de SII conforme nos alejamos del hablante, ya que los valores en las filas de detrás se reducen. En el diagrama polar, podemos ver los valores más altos de SII de un tono más amarillo y los valores bajos en tonos cada vez más azules, junto con la escala que muestra los valores máximos y mínimos obtenidos del análisis. Cabe mencionar que nuestro sistema dispone de 4 nodos de grabación, pero que se los puede mover a diferentes posiciones y volver a lanzar el código para evaluar tantas ubicaciones como se desee, por ello hemos mostrado 16 valores medidos variando 4 veces la posición de los nodos. Esta representación se lanza automáticamente en cuanto se reciben los audios y se calculan los parámetros, por lo que se puede poner el sistema a grabar y automáticamente, mostrar el análisis y almacenarlo en cuestión de segundos, con lo que esto representa de ahorro de tiempo y de sencillez a la hora de evaluar salas.

4 Conclusiones

En este trabajo se ha diseñado, prototipado y desplegado un sistema IoT de medida de SII, a partir de la norma ANSI para la evaluación acústica de salas. En la época actual y debido a la distancia social necesaria que se ha implantado en los centros docentes, es de especial interés disponer de un sistema como el diseñado para evaluar la inteligibilidad del habla en diferentes espacios. Es verdad que existen sistemas en el mercado que ofrecen análisis similares, a un precio muy elevado comparado con el sistema presentado, y con una complicación tecnológica también excesiva para un primer y rápido análisis.

Se ha reducido la complejidad del sistema al mínimo para que el usuario se despreocupe de la parte tecnológica y sólo se centre en las posiciones que desea analizar de cada sala. Así mismo, para facilitar esta tarea los nodos de nuestro sistema son totalmente autónomos en cuanto a energía se refiere y se comunican de manera inalámbrica, por lo que facilita su despliegue y ubicación en cualquier zona. El sistema diseñado implementa una visualización de los datos de una manera clara, limpia y rápida, ya que en cuestión de segundos permite ver gráficamente los resultados medidos de SII permitiendo su evaluación por el usuario.

Además, el sistema puede almacenar los datos que se deseen en un servidor en la nube y poder recuperarlos y mostrarlos en cualquier momento. Gracias a ello se pueden ensayar diferentes técnicas de mejora de inteligibilidad del habla, tanto basadas en sistemas hardware de megafonía, como software de procesado de señal, medir y comparar los resultados obtenidos de forma sencilla y rápida. Todo ello permite analizar y mejorar la inteligibilidad del habla no sólo en aulas docentes, sino en cualquier tipo de recinto donde la transmisión oral sea crítica, como puede ser un quirófano, una sala de control de un aeropuerto o una sala de espera de un hospital.

Agradecimientos

Esta investigación ha sido parcialmente financiada por el Ministerio de Innovación y Economía (ref.: BIA2016-76957-C3-1-R y BIA2016-76957-C3-2-R, cofinanciado con fondos FEDER) y por la Generalitat Valenciana, con las becas BEST/2020/117 y AEST/2020/048.

Referencias

- [1] W. C. Sabine, 1900-1915. Collected Papers on Acoustics Peninsula, Los Altos, CA.
- [2] ISO 3382-1:2009. Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces.
- [3] ISO 3382-2:2008. Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms.
- [4] ISO 3382-3:2012. Acoustics – Measurement of room acoustic parameters – Part 3: Open plan offices.
- [5] Schroeder, M.R., Integrated-Impulse Method Measuring Sound Decay without Using Impulses. J. Acoust. Soc. Am., 66 (1979) pp. 497–500
- [6] Farina, A., Simultaneous measurement of impulse response and distortion with a swept-sine technique. AES 108 th Convention, Paris, 2000 February 19-22, Preprint 5093

- [7] Vorländer, M. and Kob, M., Practical aspects of MLS measurements in building acoustics, *Applied Acoustics*, Vol. 52, No. 3–4, pp. 239–258, 1997.
- [8] Stan, G.-B., Embrechts, J.-J. and Archambeau, D., Comparison of different impulse response measurement techniques, *Journal of the Audio Engineering Society*, Vol. 50, No. 4, pp. 249–262, 2002.
- [9] Mommertz, E. and Muller, S., Measuring Impulse Responses with Digitally Pre-emphasized Pseudorandom Noise Derived from Maximum-Length Sequences, *Applied Acoustics*, Vol. 44, No. 3, pp. 195–214, 1995 }
- [10] M. Cobos, J. J. Perez, S. Felici, J. Segura and J. M. Navarro, Cumulative-sum-based localization of sound events in low-cost wire-less acoustic sensor networks. *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol.12, no.12, pp.1792-1802, 2014
- [11] A. Alexandridis and A. Mouchtaris. Multiple sound location estimation and counting in a Wireless Acoustic Sensor Network. 2015 IEEE Workshop on Applstageications of Signal Processing to Audio and Acoustics. October 18-21, 2015, New Paltz, NY.
- [12] B. Malhotra, I. Nikolaidis, J. Harms. Distributed classification of acoustic targets in wireless audio-sensor networks. *Computer Networks*.2008. DOI: :10.1016/j.comnet.2008.05.008
- [13] M. F. Duarte and Y. Hen Hu. Vehicle classification in distributed sensor networks. *J. Parallel Distrib. Comput.* 64 (2004), pp. 826838. DOI:10.1016/j.jpdc.2004.03.020
- [14] A. Pastor-Aparicio, J. Segura-Garcia, J. Lopez-Ballester, S. Felici-Castell, M. García-Pineda and J. J. Pérez-Solano, "Psychoacoustic Annoyance Implementation With Wireless Acoustic Sensor Networks for Monitoring in Smart Cities," in *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 128-136, Jan. 2020, doi: 10.1109/JIOT.2019.2946971.
- [15] T. van Waterschoot and M. Moonen, Distributed estimation and equalization of room acoustics in a Wireless Acoustic Sensor Network. 20th European Signal Processing Conference (EUSIPCO 2012). Bucharest, Romania, August 27 – 31, 2012
- [16] Larm, P., & Hongisto, V. (2006). Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index. *The Journal of the Acoustical Society of America*, 119(2), 1106–1117. <https://doi.org/10.1121/1.2146112>
- [17] Lam, Choi Ling Coriolanus., Hong Kong Polytechnic University., Department of Mechanical Engineering., *Improving the speech intelligibility in classrooms*. Dept. of Mechanical Engineering, The Hong Kong Polytechnic University; 2010.
- [18] McNeer RR, Bennett CL, Horn DB, Dudaryk R. Factors affecting acoustics and speech intelligibility in the operating room: Size matters. *Anesth Analg.* 2017;124(6):1978-1985. doi: 10.1213/ANE.0000000000002118.
- [19] Ryherd EE, Moeller M, Jr, Hsu T. Speech intelligibility in hospitals. *J Acoust Soc Am.* 2013;134(1):586-595. doi: 10.1121/1.4807034.
- [20] 1. American National Standards Institute., Acoustical Society of America., Standards Secretariat., *American national standard: Methods for calculation of the speech intelligibility index*. New York: American National Standards Institute; 1998.
- [21] J. Lopez-Ballester, A. Pastor-Aparicio, S. Felici-Castell, J. Segura-Garcia and M. Cobos, "Enabling Real-Time Computation of Psycho-Acoustic Parameters in Acoustic Sensors Using Convolutional Neural Networks," in *IEEE Sensors Journal*, vol. 20, no. 19, pp. 11429-11438, 1 Oct.1, 2020, doi: 10.1109/JSEN.2020.2995779.