REINFORCEMENT LEARNING APPLIED TO AN ACTIVE NOISE CONTROL SYSTEM IN SONIC CRYSTAL NOISE BARRIERS

David Ramírez-Solana[*1,2] **Jaime Galiana-Nieves**[2] **Javier Redondo**[2]
Agostino Marcello Mangini[1] **Maria Pia Fanti**[1]

[1] Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, Via Orabona, 4, 70125, Bari, Italia

[2] Instituto de Investigación para la Gestión Integrada de Zonas Costeras, Universitat Politècnica de València, C. Paranimf, 1., 46730 Gandia, Spain

ABSTRACT

Noise control is one of the main environmental challenges today. Transport noise is often reduced by using noise barriers which are problematic due to their lack of permeability to water and wind. In recent decades, an alternative to conventional noise barriers has been proposed based on Sonic Crystals, called Sonic Crystal Noise Barriers (SCNB). However, they present another problem due to their lack of efficiency at low frequencies.

On the other hand, cheaper technology in recent years has greatly enhanced Active Noise Control (ANC). Protection against noise is becoming more and more feasible using this type of technology. In contrast to SCNB, active noise control is more efficient at low frequencies because the sweet point is larger at these frequencies.

All this leads us to consider the combination of both ideas in this numerical approach. The aim is using a Reinforcement Learning (RL) architecture with a Double Deep Q-Network (DDQN) agent, studying the potential of incorporating ANC into these SCNB and produce permeable barriers with the best low frequency response possible.

Keywords: *Active Noise Control, Noise Barrier, Sonic crystal, FDTD, Reinforcement Learning***.**

1. INTRODUCTION

Sonic crystal noise barriers (SCNB) are a type of noise barrier that uses periodic structures to reflect or absorb sound waves of specific target frequencies [1]. These frequencies can be determined by choosing properly the topology of the scatterers or the lattice constant between them. They have shown great promise in reducing traffic noise pollution with the highlight of being permeable to light and wind, but they are often expensive to install and maintain. Active noise control (ANC) systems, on the other hand, use speakers and microphones to create an opposing sound wave that cancels out the incoming sound wave [2]. They can be more cost-effective than SCNB, but their effectiveness depends on the accuracy of the noise prediction model and the control algorithm.

Regarding control algorithms, Reinforcement Learning (RL) is a subfield of machine learning that focuses on developing algorithms that enable an agent to learn from its environment by interacting with it and receiving feedback in the form of rewards. RL algorithms such as deep Q-learning and actor-critic methods have been successfully applied to active noise control systems, allowing them to adapt to changing noise conditions and achieve better noise reduction performance, in particular, in closed spaces [3].

However, there are still several challenges to be addressed in the application of RL to ANC systems for SCNB.

[*] ***Corresponding author***: *david.ramirezsolana@poliba.it.*

On one hand, the biggest handicaps of SCNB are:

- Their poor noise reduction response in low frequencies.
- Most of the studies are made considering normal incidence that enhances its insulation in the diffractive frequency range and not considering a random incidence.
- Usually, an incident plane wave is considered as a main noise source when the source is far away enough.

On the other hand, ANC has also some drawbacks to consider:

- The main source to cancel used to be fixed and the secondary source is placed to obtain a cancellation region (sweet point) for specific cases where it is desired.
- Nearly all of approaches are made for indoor systems or enclosure cases.

Taking into account the previous considerations, this work proposes an RL procedure to find an autonomous ANC system for a SCNB, having a Finite Difference Time Domain (FDTD) simulation environment with low frequencies moving sources and outdoor propagation noise case.

2. METHODOLOGY

2.1 Main structure of Reinforcement Learning (RL)

Reinforcement Learning (RL) is a branch of machine learning that deals with sequential decision-making and takes into account artificial agents that, like biological agents, learn by interacting with their environment. The artificial agent makes use of its experience to achieve goals that are presented as a series of cumulative rewards following the scheme of Figure 1. The ability of the agent to learn appropriate behaviour, gradually modifying and acquiring new skills, and the use of trial-and-error experience are the main components of RL. The RL agent just has to be able to interact with the environment and gather information; it does not need to have comprehensive knowledge of or control over the environment [4].

2.1.1 Double Deep Q-Network Algorithm

The RL architecture chosen is a Double Deep Q-Network (DDQN) Agent. A model-free technique with off-policy based algorithm. In this case, a value-based RL agent called a DDQN agent teaches a critic (Q) to predict future rewards. Since the action space is discrete a Q-value function target critic (Q_c) is employed to improve the stability of the process. The DDQN agent takes the latest critic parameter values and updates Q_t. The agent adjusts the parameters values during training. The trained value function approximator is stored in the critic Q and the parameters are left at their tuned value after training. One issue with the DQN algorithm is that it overestimates the true rewards; the Q-values predict a bigger reward for the agent than it will really be. The DDQN algorithm advise applying a straightforward hack to address this: separating the selection of the action from the evaluation of the action. During the training phase the steps are these ones [5]:

1. Initialise the critic $Q(S,A)$ and target critic $Q_c(S,A)$ with the same values of parameters (ϕ, ϕ_c). Where S are the observations and A the actions.
2. For each episode:
    a. For the initial observation of the episode (S_{t0}) select a random action with probability ε which follows a decay function to define as a hyperparameter.
    b. Execute action (A_t) and observe the reward (R_t) and the next observation (S_{t+1}).
    c. Store the experience in a buffer (S_t, A_t, R_t, S_{t+1}).
    d. Set the next action for which the critic value function is the biggest. Also set the value function target (y_t), that can be consider the policy-decision function of the target critic (Q_c):

$$A_{max} = arg\,max\, Q\,(S_{t+1,}A_{t+1}) \quad (1)$$

$$y_t = R_t + \gamma Q_c(S_{t+1,}A_{max}), \quad (2)$$

    where $\gamma \in [0,1]$ is the discount factor an hyperparameter that represents how important are the nearest rewards over future rewards in perspective.
    e. Update critic parameters (Q) with the loss L across all stored experiences:

$$L = (y_t - Q(S_{t\,,}A_t))^2 \quad (3)$$

    f. Update the target critic (Q_c) parameters (ϕ_c) with the hyperparameter called "smooth factor" (τ) following:

$$\phi_c = \tau\phi + (1-\tau)\phi_c \qquad (4)$$

g. Update the probability ε for the next episode.
h. When S_{t+1} is a terminal observation set the value function target y_t to R_t,

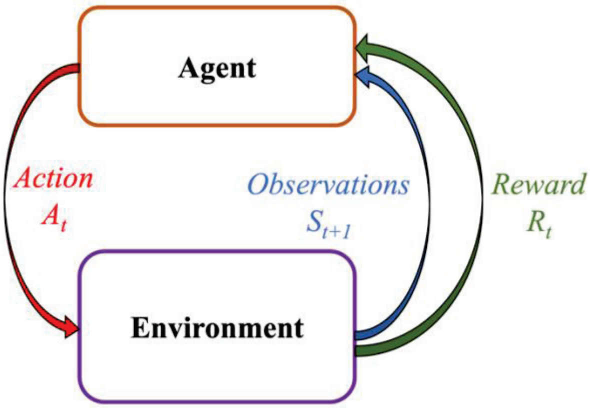

Figure 1. RL block diagram of the training process

2.2 Finite Difference Time Domain (FDTD) simulations

The finite-difference time-domain (FDTD) method is possibly one of the simplest full-wave techniques used to solve electromagnetics problems, both conceptually and in terms of its application. The FDTD method employs finite differences as approximations to the spatial and temporal derivatives appearing in Maxwell's equations. The technique was first proposed by K. Yee [6] in electromagnetics.

Maloney and Cummings [7] adapted the method to the field of acoustics using the conservation of momentum and continuity equations, which are transformed into central difference equations, obtaining update formulas for sound pressure and particle velocity.

3. ANC SYSTEM FOR SCNB WITH DDQN ARCHITECTURE

The main purpose of this work is to improve the insulation of SCNB in the low frequency range using an ANC system. So, the first decision is considering the lowest frequency in the traffic noise spectrum of international standard EN 1793-3 [8], 100 Hz. At this frequency the SNCB is almost transparent, having an insignificant insulation of only 2 or 3 dB [9]. The noise source will be a car moving at 108 Km/h in a highway section of 6 meters. This car is emitting a pure tone of 100 Hz. The barrier is placed 1 meter from the path of the noise source (car) and the evaluation point and secondary source are placed 0.5 meter from the last scatter of the SCNB. The SCNB has its main frequency insulation band at Bragg frequency = 1 kHz [10] where the traffic noise spectrum in dB(A) has more weight. So, the distant between scatterers called "lattice constant" is 0.17 m and the total width of the barrier is 0.47 m.

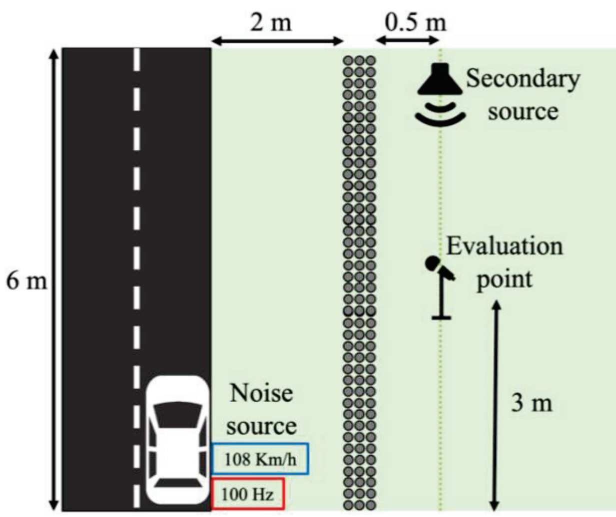

Figure 2. Scheme of the environment with the primary source at the left of the SCNB and the ANC system in the right part

3.1 FDTD Environment

The FDTD environment is equivalent to the one represented in Figure 2 but adding Perfectly Matched Layers (PML) at the boundaries to prevent unwanted reflections [11].
In the evaluation point the result of the loudspeaker cancellation signal and the car noise is obtained, looking for a noise cancellation due to the ANC principle. The FDTD simulation is run in a different function that the agent calls each time wants to act into the environment. So, this environment has either inputs as an agent actions or the time duration of the steps, and also outputs as observations. With time domain simulations techniques as FDTD is possible to interact at each episode easier than with frequency-domain simulation techniques.

3.2 Observation

The main purpose of the RL architecture is to improve the performance of the SCNB with the ANC. So, assuming the insulation produced only by the barrier when the ANC is not working as an acoustic pressure level to reduce, this is the initial observation of the process. With this initial observation, after the actions are executed a second pressure is registered and the Insertion Loss of the ANC system (IL_{ANC}) is calculated as follows:

$$IL_{ANC} = 10 \times log_{10}\left(\frac{\sum_t^{150} p(t)_b}{\sum_t^{150} p(t)_{ANC}}\right)[dB], \tag{5}$$

where $p(t)_b$ is the acoustic noise pressure with only the SCNB acting as a noise mitigation device and $p(t)_{ANC}$ the acoustic noise pressure obtained after the secondary source is emitting a signal to cancel the primary noise. The end of each episode is defined by the last time step, with a value of 200 ms the time that the car needs to pass from the bottom to the top of the environment. Following the fast mechanism of usual sound meters, that gives a value each 125 ms, and letting the RL system to modify the action during the car is passing a value of 100 ms for each step is defined.

3.3 Actions

The actions are discrete, since it is a FDTD simulation environment with discrete mesh, as well as a DDQN architecture. Both actions have the goal of produces the most suitable signal that can cancel the noise emitted from the primary source (car), the first one (action 1) defines the amplitude of the source and the second one (action 2) the phase of the signal. When the phases of the signals are opposed, the acoustic wave is cancelled, and a "sweet point" is produced thanks to the ANC main principle. The signal of the secondary source (*S2*), follows the next equation:

$$S2 = -Action_1 \times sin(2\pi ft + Action_2), \tag{6}$$

with f = 100 Hz and t a vector of the time that each step of the environment takes to evaluate the observation. The time vector is defined as 100 ms, ,then the observation is passed to the next setp and each episode has 2 steps and 2 different actions to interact with the RL environment. So, every car passing is considered as an episode where 2 actions can be executed with 2 different observation and rewards.

3.4 Reward

The agent performs its behaviour at each call to the step function (defined as 100 ms) and receives rewards regarding on how those actions affect the environment according to the observations. The reward can be thought of as the environment's feedback, indicating whether the agent's activities were successful or unsuccessful. The agent needs to modify some parameters of the signal radiated from the secondary source, to find the best noise cancellation result in the evaluation point. Considering the observation previously described in section 3.2, the Reward function analyse this IL_{ANC} and according to how much is improving the insulation is proportionally setting its value:

$$Reward = \begin{cases} IL_{ANC,} & IL_{ANC} > 0 \\ 0, & IL_{ANC} < 0 \end{cases}, \tag{7}$$

where there is no reward if the ANC system is augmenting the noise pressure level (IL_{ANC})<0, and on the other case the exact level of dBs that increase the initial observation using the current observation of the step, according to equation (5). By doing this, even if the reward is very small, it will be always an improvement with respect not having the ANC system and have only the SCNB.

4. RESULTS

4.1 Training of the DDQN Agent

During the training, the DDQN agent updates its critic model at each episode. The hyperparameters have been chosen according to **Table 1.** Where ε greedy probability distribution over the episodes guarantees a wide exploration of the parameters over the number of episodes and the optimizing hyperparameters are giving more importance to the future rewards according to the saving values, as the high discount factor (γ) ensures.

Table 1. Training hyperparameters

Parameter	Value
Critic Optimizer type	adam
Critic Optimizer learn rate	0.01
Agent Discount factor (γ)	0.99
Agent Batch size	64
Agent buffer length	1000
Agent smooth factor (τ)	0.001
Initial epsilon greedy (ε_{ini})	1
Epsilon greedy decay (ε_{decay})	0.01
Minimum epsilon greedy (ε_{min})	0.01

In Fig. 3 the training process of the system is represented with only 500 episodes, since this study case is very simple with only one frequency and only one velocity of the moving source. The convergence is found in episode 180 when the agent is already trained. The cumulative rewards of the episode and the average reward are plotted, showing a smoothy tendence in the averaged case. Also, the maximum cumulative reward is plotted.

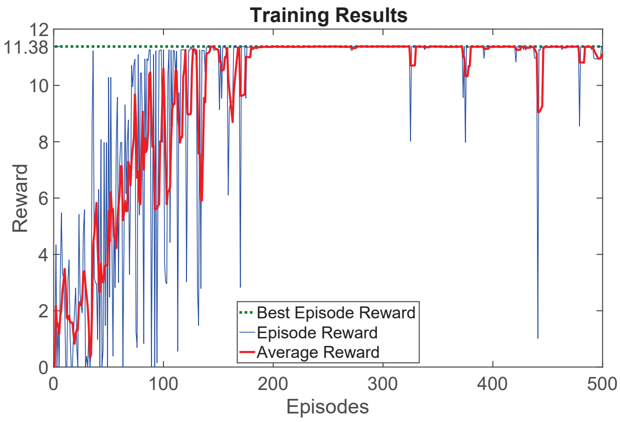

Figure 3. -Training results of the DDQN Agent

Considering that each episode has two steps, and the cumulative reward of each episode is the sum of the step's rewards we can assume that the averaged value of insulation in dBs could be the half of the episode reward. So, taking the two best possible actions, an averaged improvement of 5.7 dB can be obtained.

4.2 Deployment of the environment

After the training, the agent had found the best value of actions that satisfy the maximum IL improvement in the two steps of each episode. So that is why all the simulations run with the DDQN agent have constant values as Fig. 4 shows. This is the optimal value of the actions to produce the bigger noise cancellation at the evaluation point. With 8.39 dB in the first 100 ms and 3.40 dB in the second 100 ms. When the agent is not trained, the reward is always equal to zero since the secondary source is contributing in negative way to mitigate the sound produced by the car.

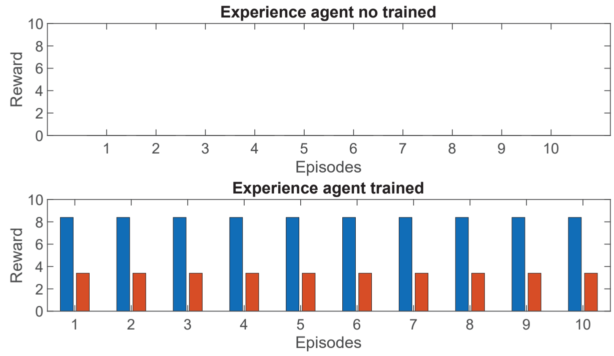

Figure 4. Simulations with the DDQN agent without training knowledge (up) and after when it is trained (down)

4.3 ANC Analysis

The resulting insulation of the SCNB considers both phenomena; SNCB and ANC to mitigate the noise of the primary moving source (car). In the next figure, is presented the reset function simulation to obtain the insulation produced by the SCNB without the ANC system working.

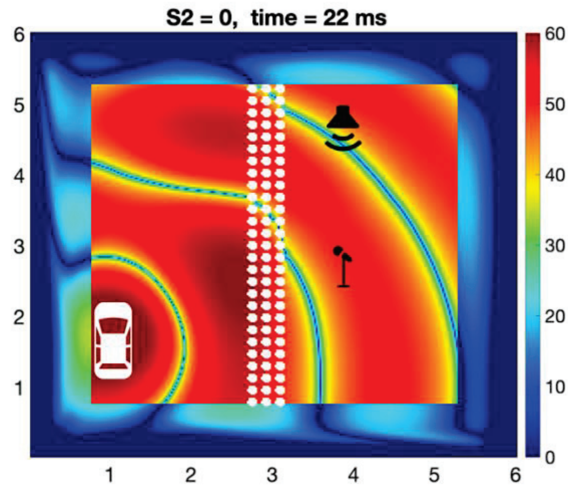

Figure 5. FDTD simulation with only the primary source emitting.

When the ANC system is working, the actions modify the signal of the secondary source (*S2)* so one of the best cases is represented in Fig. 6 when the first step's IL_{ANC} = 8.39 dB during the first step period (100 ms) is obtained.

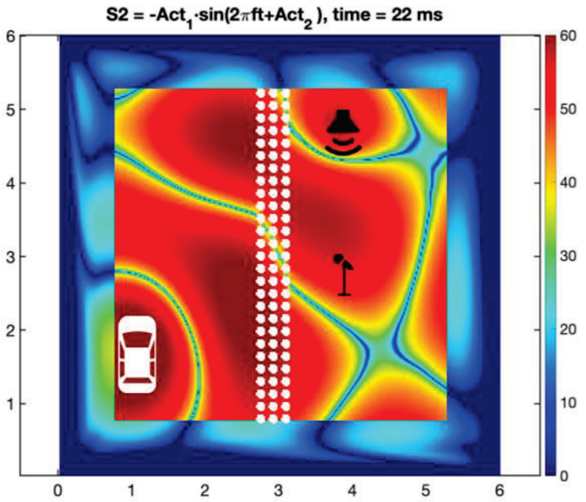

Figure 6. – FDTD simulation with the primary and secondary source emitting

4.4 Insertion Loss results

In the introduction it has been said that SCNB are usually studied under a normal incidence condition of the noise source, but according to the case of this study, a random incidence average for the Insertion Loss (IL) following the next equation is consider:

$$IL_{rand} = -10 \cdot log_{10}\left(\frac{\int_{0°}^{\theta_{lim}} \frac{p_b(f,\theta)}{p_i(f,\theta)} cos\,(\theta) d\theta}{\int_{0°}^{\theta_{lim}} cos\,(\theta) d\theta}\right), \qquad (6)$$

where $p_b(f,\theta)$ and $p_i(f,\theta)$ are the acoustic pressures measured after placing the barrier and without placing it respectively, for a specific frequency f, and incident angle θ. The angle limit is θ_{lim}= 85° since according to several studies that said that after that angle there is no incident acoustic energy arriving to the barrier [12] [13]. Also, the $cos(\theta)$ term assures that the contribution of the angles above the limit are almost neglectable. In Fig. 4 an example of the IL averaged of a SCNB with 3 rows is presented in blue colour, further explanation can be found in [14], the barrier is analogous to the one employed in this study. The best improvement of 5.7 dB is added to the 2.16 dB at 100 Hz, resulting in a IL value of 7.86 dB. Another interesting frequency to improve would be 271 Hz where the value of the IL is almost zero, but the higher the frequency is, the worse performance the ANC system has. This is due to the reason that the sweet point becomes smaller and more difficult to control.

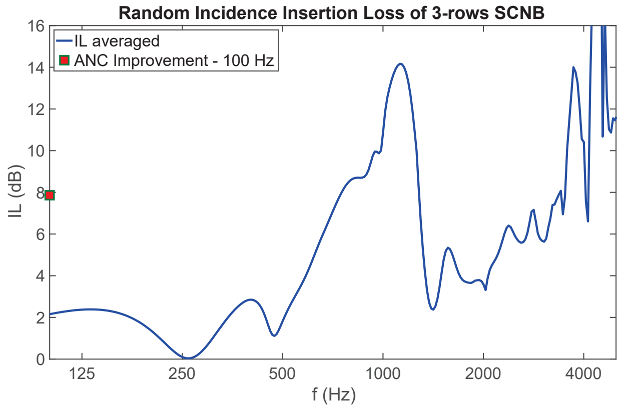

Figure 7. Insertion Loss with random incidence average of 3-rows SCNB and the ANC improvement at 100 Hz.

5. CONCLUSION

One of the highlights of this study is the procedure to train an autonomously SCNB active barrier throw a time domain simulation as FDTD and with a RL architecture.
In the low-frequency range, the test evaluates a specific case of 100 Hz, the lower frequency that European Standards considers.
Reaching a value of 7.86 dB, bigger than the triple of that frequency without the ANC system, an improvement to our barrier brings the potential of applying this technique in more frequencies and cases.

As future research a LMS algorithm can be applied to obtain a better cancellation with the ANC system. Also, random frequencies and velocities can be settled to train the system in order to make it more reliable to real cases and also different car velocities.

6. REFERENCES

[1] J. Sánchez-Pérez, C. Rubio, R. Martinez-Sala, R. Sanchez-Grandia and V. Gomez, "Acoustic barriers based on periodic arrays of scatterers," *Applied Physics Letters,* vol. 81, no. 27, pp. 5240-5242, 2002.

[2] J. Guo and J. Pan, “Increasing the insertion loss of noise barriers using an active-control system,” *The Journal of the Acoustical Society of America,* vol. 104, no. 6, pp. 3408-3416, 1998.

[3] B. Raeisy, S. G. Haghighi and A. Safavi, “Active noise control system via multi-agent credit assignment,” *Journal of Intelligent & Fuzzy Systems,* vol. 26, pp. 1051-1063, 2014.

[4] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare and J. Pineau, “An introduction to deep reinforcement learning,” *Foundations and Trends in Machine Learning,* vol. 11, pp. 219-354, 2018.

[5] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction., Cambridge: The MIT Press, 1992.

[6] K. Yee, “Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media,” *EEE Transactions on Antennas and Propagation,* vol. 14, no. 3, pp. 302-307, 1966.

[7] J. Maloney and K. Cummings, “Adaptation of FDTD techniques to acoustic modelling,” *1th Annu. Rev. Prog. Applied Computational Electromagnetics,* vol. 2, p. 724, 1995.

[8] European Committee for Standardization. EN 1793-3:1997, *Road traffic noise reducing devices – test method for determining the acoustic performance – Part 3: Normalized traffic noise spectrum,* 1997.

[9] S. M. Dimitrijevic, V. M. García-Chocano, F. Cervera, E. Roth and J. Sánchez-Dehesa, “Sound Insulation and Reflection Properties of Sonic Crystal Barrier Based on Micro-Perforated Cylinders,” *Materials,* vol. 12, no. 2806, 2019.

[10] Kittel, C., Introduction to Solid State Physics, 8a Ed., 2004.

[11] J. Berenguer, “A perfectly matched layer for the absorption of electromagnetic waves,” *J. Comput. Phys.,* vol. 114, pp. 185-200, 1994.

[12] S. B., “Prediction methods for the sound transmission of building elements,” *Noise Control Eng. J.,* vol. 11, p. 5363, 1978.

[13] H. Kang, J. Kim and K. H., “Prediction of sound transmission loss through multilayered panels by using Gaussian distribution of directional incident energy,” *The Journal of Acoustical Society of America,* vol. 107, pp. 1413-1420, 2000.

[14] D. Ramírez-Solana, J. Galiana-Nieves, J. Redondo, A. M. Mangini and M. P. Fanti, “Análisis numérico del aislamiento producido por barreras basadas en cristales de sonido en incidencia oblicua,” in *53º CONGRESO ESPAÑOL DE ACÚSTICA XII CONGRESO IBÉRICO DE ACÚSTICA*, Elche (Spain), 2022.