2020 International Conference on Computational Science and Computational Intelligence (CSCI)

A Conceptual Model for real-time Binaural-Room Impulse Responses generation using ANNs in Virtual Environments: State of the Art

Daniel A. Sanaguano
Departamento de Informática y Ciencias de la Computación
Escuela Politecnica Nacional
Quito, Ecuador
Email: daniel.sanaguano@epn.edu.ec

Jose F. Lucio-Naranjo
Departamento de Informática y Ciencias de la Computación
Escuela Politecnica Nacional
Quito, Ecuador
Email: daniel.sanaguano@epn.edu.ec

Roberto A. Tenenbaum
Programa de Pós-Graduação em Engenharia Civil
Universidade Federal de Santa Maria
Santa Maria, Brazil
https://orcid.org/0000-0002-5268-3849

***Abstract*—This work aims to give an overview of Artificial Neural Networks (ANN) approaches applied for BIRs generation published in the literature and to expose gaps in the academic research. The literature review shows that several successful studies are using ANNs approaches for BIRs generation with a reduction in the computational effort by up to 90% with respect to the Traditional Method. Nevertheless, these approaches are bounded by a fixed pair of a sound-source and binaural-receptor, meaning that they do not take into account dynamic variations in the position of the receptor. In this sense, this work also introduces a conceptual model for a real-time BIRs generator that considers a moving binaural-receptor using a set of Artificial Neural Networks.**

***Keywords*-Real-time Auralization; Artificial Neural Networks; Binaural-Room Impulse Responses; Acoustic Virtual Reality.**

I. Introduction

Acoustic Virtual Reality (AVR) generates spatial audio from a virtual environment (such as rooms, auditoriums, halls, etc.) that simulates the acoustical sensation that a person could have if present in the environment. This is done by using complex algorithms of computational modeling that include the sound wave propagation and an auralization technique [1]–[4]. The first studies in auralization appeared in the 1990s [5]–[7]. Subsequently, [3], [8]–[10] present guidelines in order to implement auralization using Binaural-Room Impulse Responses (BIR) generation. This procedure will hereinafter be referred to as the Traditional Method (TM). In the following years, the studies for BIRs generation increased, mainly because of the development of powerful processing machines and more efficient algorithms partially solving the high computational requirements that convolution techniques demand (e.g. [11], [12]).

An Acoustic Computational Simulation requires numerical modelling of the propagation of a sound wave from the sound source until it reaches the acoustic receptor. However, this propagation is a complex process due to physical phenomena such as reflection, scattering, and diffraction. Also, the composition and geometry of the contour materials increase the complexity. Many sound wave fronts classified according to their arrival to the receptor must be carefully modeled. This must be conducted for every pair of an sound-source and an acoustic-receptor present in the virtual environment. Each pair constitutes an acoustic transmission system and its characteristics are fully represented by the Room Impulse Response (RIR), with [13] being the pioneering work. To make matters even more complicated, in humans, the sound wave is also modified by anthropometric features before reaching the two receptors (left and right ear). Namely, an acoustic representation in humans requires two functions, instead of one, known as the Binaural-Room Impulse Responses (BIRs). [3] proposed a model to generate the BIRs using a pair of Head-Related Impulses Responses (HRIRs) or its frequency representation, the so called Head-Related Transfer Functions (HRTFs). In short, the BIR process consists of a sum of all the convolution products between the HRIRs with the arriving wave-fronts [3]. In other words, the HRIRs are embodied in the BIRs, and these depend on the direction of incidence (i.e. azimuth and elevation angles) of the wave fronts and the anthropometric characteristics (e.g. size of head, torso, ears drums, pinna, etc.) [1]. To recreate a spatial audio perception without undesirable effects [14], any incidence direction of the wave-front is expected to have an HRTF, making mandatory an interpolation procedure in order to compensate for the discrete nature of the measurements present in existing HRTF databases [15], [16].

In [1], the author describes two main phases with a high computational cost for any AVR generation using the Room Acoustic Simulation approach. The first phase is related to the wave propagation in the given environment. This consists of calculating all wave front representations (acoustic rays) that arrive at each receiver. The second phase is the auralization, which comprises the computational process of the BIRs for each sound source receptor pair.

978-1-7281-7624-6/20/$31.00 ©2020 IEEE
DOI 10.1109/CSCI51800.2020.00102

According to [3], [17], a method of Geometric Acoustic (GA) that uses Image Source (IS) is a reliable approach for Acoustic Computational Simulation. Nevertheless, it requires high computational processing for high precision simulation. For instance, for a solid simulation, each sound source in the AVR must produce at least 10^5 acoustic rays, according to [18], [19].

In the 2000s, due to an increase in computational power, Machine Learning methods became popular, and ANNs more recently. These methods are data-driven, which means they do not rely on actual knowledge of the system that transforms inputs into expected outputs. Additionally, Machine Learning fixed well-known problems such as non-linearity predictions, generalization, and noise tolerance with a less complex model and efficient computational performance, once the training procedure has ended [18]. Thanks to these benefits, [20] carried out a new model to generate HRTFs using Artificial Neural Network (ANN), further, [1], [18], [21] implemented a BIR generation model using HRIRs learned by ANNs as an alternative to TM, obtaining a reduction in the computation effort.

In spite of the good results of the aforementioned studies, an efficient HRIRs generation is not enough to guarantee a precise BIRs generation in real time, and therefore, a reliable auralization. This is because the ANN procedures only deal with the modified HRIRs generation (i.e. the convolution between the incident wave-fronts of an impulsive signal and the corresponding HRIR regards to incident angle). This procedure must be repeated for each wave-front that appears in the simulation, which is evidently computationally expensive. To make matters even more complicated, if a position shift occurs in the receptor or even in its orientations with regards of the sound source, a new BIR must be generated from scratch. In this sense, this works aims to identify academic research gaps through a systematic literature review in order to justify the proposal of a new model for a reliable real-time BIR generation using a set of ANNs.

This work is presented as follows. Section II describes the whole search process. This section comprises the Research questions, the search strategy applied, and the process of studies selection. Section III presents the reviewed studies as results of the research process, and the main identified issues. Section IV describes in detail the real-time BIR generation model using ANNs. Section V evidences the constraints of the model. Section VI emphasizes on the detected gaps and suggests future researches.

II. Research Method

This study presents the state of the art of the literature related to Artificial Neural Networks approaches for real-time BIRs generation in Virtual Environments. In this sense, the literature review process was based partially on [22]. This section is divided into Planning and Searching, Studies' Selection, and Data Analysis.

A. Planning and Searching

This subsection introduces the Research Questions (RQs) that intends to answer the main objectives of this study. It also describes the search process, and the inclusion and exclusion criteria, as it is shown below.

- **RQ1.** What are the approaches used for the generation of BIR (with/without real-time) as an alternative of TM?
- **RQ2.** What is the appropriate ANN architecture and spatial distribution of an ANN set that provide an accuracy (>80%) in training and validation step with less computational cost in estimating the BIRs?
- **RQ3.** What is an appropriate ANN architecture to spatially interpolate a function and which is the performance measure to be consider?
- **RQ4.** What are the subjective validation tests used in the literature to guarantee a reliable auralization?

The Search process enclosed three combinations of terms related to Artificial Neural Networks (ANNs), BIR generation, spatial distribution, real-time and computational cost. In order to minimize the research bias, we only selected peer-review journals from the scientific databases ScienceDirect/Elsevier, Scopus, Springer, and IEEE-Xplore. The search criteria also included only studies written in English and available works since 2015. The literature review was carried out between January-July 2020.

B. Study Selection

The search process found 414 studies. To reduce redundancy and research bias, two filters were applied. 1) The Wipe filter dropped off duplicates journals and unrelated topics. This resulted in 79 journals. 2) The Deep Filter eliminated 57 studies that did not answer the research questions aforementioned, remaining 22. By snowballing technique [22], 8 works were added, obtaining 30 journals in total.

C. Data Extraction

In order to classify the papers into meaningful structures and thus tackle the research questions, a set of questions must be answered after a fulled-review text. Table II shows the questions carried out across 30 journals. The outcomes are described in the next section.

Table I
RESULTS OF THE SEARCH PROCESS DONE BETWEEN JANUARY AND JULY 2020. ONLY WRITTEN ENGLISH PEER-VIEW JOURNALS WERE SELECTED SINCE 2015.

Topic	Searched Term	Scopus	Springer	Science Direct	IEEE Xplore	Total
Acoustic Virtual Environment	Binaural-Room Impulse Response* AND Artificial Neural Network **	23	2	5	1	31
	Artificial Neural Network AND Acoustic Virtual Reality	5	0	1	1	7
Spatial-Interpolation	Grid-Spatial Interpolation AND Artificial Neural Network	60	90	141	1	292
	Set of Artificial Neural Networks AND Interpolation	3	0	0	21	24
Auralization	Auralization AND Computational Cost	16	21	19	0	56
	Head-related Impulse Response AND Auralization AND Artificial Neural Network	3	0	0	1	4
					Total	414

Table II
DATA EXTRACTION QUESTIONS DONE FOR EACH CANDIDATE STUDY.

Feature	Code	Question
BIR generation Method	EQ1	Was the BIR generation based on ANN?
	EQ2	Was there a real-time BIR generation?
	EQ3	Was there a computational cost measure?
ANN approaches	EQ4	What were the inputs and outputs of the ANN?
	EQ5	Which ANN architecture was used ?
	EQ6	Did it use a grid or a set of ANN?
	EQ7	What percentage of data were used for training and testing?
Experimental Validation	EQ8	Which statistic validation techniques were used ?
	EQ9	How many datasets were compared ?
Performance measures	EQ10	Which metrics were used to assess the model's performance?
Interpolation Methods	EQ11	How was the spatial distribution divided?
	EQ12	Was a time-frequency domain function interpolated in space?

III. LITERATURE REVIEW RESULTS

This section arranges chronologically and points out the most relevant works according to the methodology defined above. The aim was to identify the main drawbacks of real-time BIR generation and spatial interpolation constraints. The first characteristic encompasses the BIR generation for a Virtual Acoustic Environment. As mentioned, BIR generation is a key part of auralization in room acoustic simulation. Some well-known studies encountered were [4], [5], [7]–[10], [23]. [8] carried out a model of BIR generation for Room Acoustic that comprises two phases: 1) Simulate the wave propagation from the sound source across the room, until reaching the receiver, considering reflection and scattering with internal surfaces. A hybrid model was proposed by [8] based on image source for specular reflections and the stochastic ray-tracing for scatter phenomena; 2) Generate the BIR for a static point in space by summing up the convolution products between the incident wavefront representation and the corresponding Head-Related Impulse Response (HRIR) for incident direction. This procedure can produce a sound with immersive characteristics in the given environment. In this sense, several works have adopted this model along with computer-aided design (CAD) in order to render faithful simulation [24]–[26].

However, to achieve a reliable simulation, at least 10^5 rays per each pair of sound-source and receiver are necessary for a given scenario according to [3], [19]. This increases the computational cost and also makes a real-time BIR generation an even more complex task. Additionally, to validate the real-time BIR in humans, subjective evaluations and experiments must be conducted.

Additionally, [27] carried out an attempt to implement a real-time system for an Auditory Virtual Environments using the traditional approach. It is worth noting, that TM process must be repeated for each variation in the receptor position considering only one fixed sound-source.

Machine Learning became popular because it fixed problems related to non-linearity predictions, generalization, and noise tolerance. This conveys a reduction of model complexity and computational costs after the training step. Thanks to these benefits, several studies have been carried out in the acoustic field including sound source localization [28]–[30], speech recognition [31]. There are several studies in the Acoustic Virtual Environment field such as HRIR's estimation and interpolation [32]–[34], personalizing HRIRs [35]–[38], Auralization [1], [18], [21], BIR's classification [39], [40], and others [17], [25].

The extracted features selected in the methodology such as the ANN approaches, performance measurements, and experimental validation for BIR generation were analyzed together. In [34], the authors aimed to speed up the implementation of Acoustic Virtual Reality using Artificial Neural Network (ANN) that generates BIR through HRIR at each point of the space. The model comprised two Multi-layer Perceptron Feed-Forward back propagation ANN, one per ear. ANN's architecture contained an input vector with the 23 features (21 anthropometric measures, the azimuth angle and elevation angle), one hidden layer with 50 neurons, and an output vector with 512 HRIR's samples. After the training, the Mean Square Error (MSE) achieved was between 3-5 %. In this sense, [38] performed another approach in estimating a personalized HRIR. The authors defined two sets of 50 feed-forward ANNs (i.e. one per ear) to cover the media plane of binaural receptor. For each ANN the input vector dimensionality was reduced to seven most relevant anthropometric measures. The best performance was achieved with one hidden layer with 18 neurons. The first six principal components of Spectral Distortion (SD) were selected as outputs that were used later to estimate the HRIR. An approach for HRIR interpolation using a Multi-Layer Perceptron (MLP) and Super Cascade Feed-Forward ANNs over the whole auditory reception area was proposed by [33]. Three signals sets were established as inputs, but it were not detailed by the authors. The output vector was an array of 128 frequency bins of HRTF. The first input set considered x,y,z coordinates; the second only azimuth and elevation; and the last set combined the elements from the first and second set. The best performance for MLP was reached with two hidden layers with 64 and 32 neurons, respectively. The Cascade Feed-Forward only used one hidden layer with 64 neurons.

[18] posed a new approach of BIR generation for auralization using a set of ANNs. This approach reconstructs the HRIRs by means of the octave bands spectral modification, and spatial interpolation. By doing so, the author proposed to divide the auditory space, represented by sphere around the head, into 1898 sections. After that, in each section a Multi-Layer feed-forward ANN was trained with: two hidden layers, using a input vector with 11 dimensions (9 for spectral octave bands, azimuth angle, and elevation), and output vector with 128 samples of HRIRs. The best performance was reached with 7 and 3 neurons in the hidden layers, respectively. To measure the performance of estimated HRIR the Mean Square Error (MSE) and Octave-MSE (OMSE) were used. Finally, a BIR comparison between the ANN approach and TM was done. The authors' approach showed a reduction of up to the 60 % in the cost computational with respect to TM .

[21] also proposed a set of ANNs as an alternative to TM for the generation of BIRs. The auditory space around the user was also pictured as a sphere divided into 64,442 sections, matching with the numbers of HRIRs measurements in the "Fabian" dummy head data-set. In each section a Radial Basis Function (RBF) ANN was trained with 11 inputs (i.e. 9-octave bands spectrum, azimuth, and elevation angle), one hidden layer, and 128 temporal samples of HRIR as outputs. The author asserted that the best cost-benefit ratio was achieved with 5 neurons in the hidden layer, reducing the computational cost by approximately $\approx 90\%$ with respect to TM. Like [18], the BIR generation was implemented using the TM.

The last feature intends to find well-known mathematical methods for spatial interpolation. The literature process led us to infer that statistical models are suited for this task due to better generalization and solving complex non-linearity relations between inputs and outputs [41]. The two main classes of statistical models for spatial interpolation are deterministic (e.g. Inverse Distance Weight (IDW), Linear Regression (LM), Radial Basis Function (RBF)), and geo-statistical techniques (e.g. Ordinal Kriging) [41], [42]; and lately, the introduction of Machine Learning techniques [43]. The studies mentioned conclude that the Kriging approach provides better performance in a wide coverage area, but for a narrow area or low spatial resolution, there is no evidence which method is better.

In summary, the previous studies have provided fruitful advances towards an efficient BIR generation. However, those efforts do not guarantee a fast and reliable BIR's computation for real-time purposes. This is because a significant variation in the receiver position requires a different BIR estimation. This means that the whole ANN must be trained each time the user changes its position.

IV. Proposed Methodology

This study presents a conceptual model for reliable BIRs generation in real-time. This will be done by interpolating the BIR in the space using a set of previously trained ANNs. Figure 1 summarizes the proposed conceptual model. To obtain accurate results, first its required a sufficient number of reliable BIRs distributed over the consider scenario. For this reason, and also for comparison purposes, two sets of BIRs will be obtained. The first set of BIR will be gathered through metrological measurements; and the second set will be obtained by means of Room Acoustic Simulation. Both will be conducted for a previously determined environment (e.g. a classroom). The points where the BIRs will be generated must be spatially distributed into a 3-D grid, in the sections where a human acoustic receptor usually can be found (standing position and sitting position with height variations). The generated BIR databases will consider a one-fixed sound-source, placed in a given position; and a binaural-receiver (dummy-head), located at each

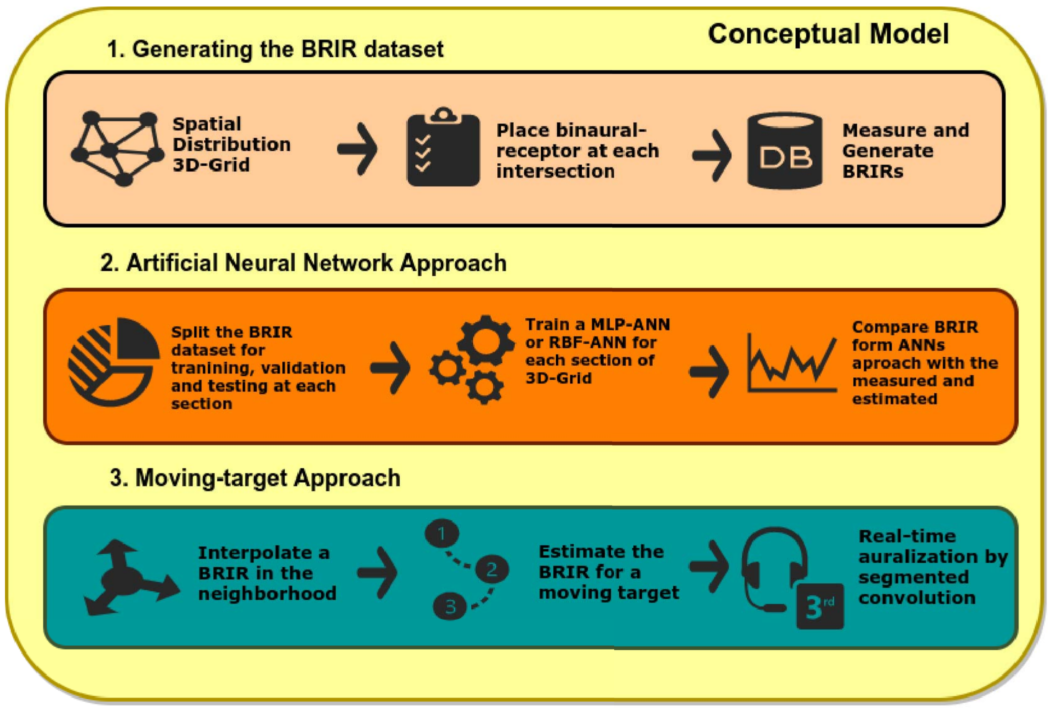

Figure 1. Conceptual Model for real-time BIR generation through a set of ANNs. Source author.

intersection of the 3D-grid, using several orientations. The measure BIRs database and the simulated BIRs database will use the same configuration.

After that, the 3D-grid will be divided into sections in order to reduce the problem complexity. Then, for each section an MLP-ANN and RBF-ANN will be trained with positional coordinates and orientation as inputs and BIRs as targets. The appropriate ANN architecture will be defined by a heuristic process. The expected output of the properly trained ANN will be an interpolated BIR with a significantly less computational cost and an adequate accuracy. Finally, real-time auralization will be achieved by the segmented convolution between the interpolated BIR and an anechoic sound signal.

V. Discussion and Limitations

Despite there being several studies in the literature related to BIRs generation using the ANN methods leading to a computational saving time of up 90 % with respect to TM, those do not take into account the change in the BIR if a position shift occurs in the receiver. This would demand a new wave propagation simulation in order to update the BIR, making an accurate real-time generation almost impossible. This research proposal intends to cover those problems, by designing a different conceptual model for efficient generation of BIR using ANNs set. However, this proposal contains two main challenges.The first is to provide an efficient and adequate design of the ANN set. The Second is to generate reliable BIRs measurements to generate an appropriate database so much for the training and validation of the ANN.

VI. Conclusion and Future Work

According to the literature review, the BIR generation using the ANNs method for the fixed binaural receptor is a recent approach in the AVR field with promising results. As mentioned, the ANNs approaches reduce significantly the number of operations, by between 60% and 90%, to generate reliable BIRs. Nevertheless, those methods only deal with the auralization for a static binaural-receptor. This means that if a position shift occurs in the binaural receptor inside the virtual environment, or even a change in the relative orientation towards the sound source, the BIR is different. Therefore, an approach for BIR real-time generation must deal with a more complex model and must run with less computational effort. However, an approach based on these demands was completely absent in the conducted literature review.

The proposed conceptual model presents a set of ANN for real-time BIR generation based on research gaps identified by the literature review. Accordingly, the conceptual model mainly tackles the problem of a reliable and fast BIR generation for a moving target by means of spatial interpolation of the BIR using a set of ANN. Nonetheless, to validate an Acoustic Virtual Reality experiment it is fundamental to conduct a subjective test, this will be part of future research.

REFERENCES

[1] R. A. Tenenbaum, F. O. Taminato, V. S. G. Melo, and J. C. B. Torres, "Auralization generated by modeling HRIRs with artificial neural networks and its validation using articulation tests," *Applied Acoustics*, vol. 130, pp. 260–269, 2018.

[2] V. Pulkki and U. P. Svensson, "Machine-learning-based estimation and rendering of scattering in virtual reality," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2664–2676, 2019. [Online]. Available: http://dx.doi.org/10.1121/1.5095875

[3] M. Vorländer, "Virtual acoustics," pp. 307–318, 2014.

[4] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization-an overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861–875, 1993.

[5] H. Lehnert and J. Blauert, "Principles of binaural room simulation," *Applied Acoustics*, vol. 36, no. 3-4, pp. 259–291, 1992.

[6] J.-P. Vian and J. Martin, "Binaural room acoustics simulation: Practical uses and applications," *Applied Acoustics*, vol. 36, no. 3-4, pp. 293–305, 1992.

[7] N. Xiang and J. Blauert, "Binaural scale modelling for auralisation and prediction of acoustics in auditoria," *Applied Acoustics*, vol. 38, no. 2-4, pp. 267–290, 1993.

[8] M. Vorländer, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Science & Business Media, 2007.

[9] F. Wefers, D. Schröder, S. Pelzer, and M. Vorländer, "Real-time filtering for interactive virtual acoustic prototyping," in *Proc. Euronoise*, 2009.

[10] M. Vorländer, D. Schröder, S. Pelzer, and F. Wefers, "Virtual reality for architectural acoustics," *Journal of Building Performance Simulation*, vol. 8, no. 1, pp. 15–25, 2015.

[11] R. Sottek, "Virtual binaural auralisation of product sound quality importance and application in practice," *Proc. EURONOISE Naples*, 2003.

[12] J. C. B. Torres, M. R. Petraglia, and R. A. Tenenbaum, "An efficient wavelet-based hrtf model for auralization," *Acta Acustica united with Acustica*, vol. 90, no. 1, pp. 108–120, 2004.

[13] M. Schroeder, "Novel uses of digital computers in room acoustics," *The Journal of the Acoustical Society of America*, vol. 33, no. 11, pp. 1669–1669, 1961.

[14] G. Wersényi, "Effect of emulated head-tracking for reducing localization errors in virtual audio simulation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 2, pp. 247–252, 2009.

[15] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The cipic hrtf database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*. IEEE, 2001, pp. 99–102.

[16] B. Xie, X. Zhong, D. Rao, and Z. Liang, "Head-related transfer function database and its analyses," *Science in China Series G: Physics, Mechanics and Astronomy*, vol. 50, no. 3, pp. 267–280, 2007.

[17] C. Schissler, C. Loftin, and D. Manocha, "Acoustic Classification and Optimization for Multi-Modal Rendering of Real-World Scenes," *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, vol. 24, no. 3, pp. 1246–1259, mar 2018.

[18] J. F. L. Naranjo and R. A. Tenenbaum, "Intelig^encia computacional aplicada na geracao de respostas impulsivas bi-auriculares e em aurilizacao de salas," Ph.D. dissertation, Universidade do Estado do Rio de Janeiro, 2014.

[19] I. Bork, "A comparison of room simulation software-the 2nd round robin on room acoustical computer simulation," *Acta Acustica united with Acustica*, vol. 86, no. 6, pp. 943–956, 2000.

[20] Z. Haraszy, D. Ianchis, and V. Tiponut, "Generation of the head related transfer functions using artificial neural networks," in *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, no. 13. World Scientific and Engineering Academy and Society, 2009.

[21] R. A. Tenenbaum, F. O. Taminato, and V. S. Melo, "Fast auralization using radial basis functions type of artificial neural network techniques," *Applied Acoustics*, vol. 157, p. 106993, 2020. [Online]. Available: https://doi.org/10.1016/j.apacoust.2019.07.041

[22] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," in *Engineering*, 2007, vol. 2, p. 1051.

[23] H. Kuttruff, *Room acoustics*. Crc Press, 2016.

[24] D. Schröder, S. Pelzer, M. Vetterli, and M. Vorländer, "Through the hourglass: A faithful audiovisual reconstruction of the old montreux casino," *Acoustics Australia*, vol. 43, no. 1, pp. 49–57, 2015. [Online]. Available: http://dx.doi.org/10.1007/s40857-015-0009-4

[25] Á. Török, D. Mestre, F. Honbolygó, P. Mallet, J.-M. Pergandi, and V. Csépe, "It sounds real when you see it. Realistic sound source simulation in multimodal virtual environments," *Journal on Multimodal User Interfaces*, no. 4, pp. 323–331, 2015.

[26] B. N. Postma and B. F. Katz, "Creation and calibration method of acoustical models for historic virtual reality auralizations," *Virtual Reality*, vol. 19, no. 3-4, pp. 161–180, 2015.

[27] F. C. Tommasini, O. A. Ramos, M. X. Hüg, and S. P. Ferreyra, "A Computational Model to Implement Binaural Synthesis in a Hard Real-Time Auditory Virtual Environment," *Acoustics Australia*, vol. 47, no. 1, pp. 51–66, 2019. [Online]. Available: https://doi.org/10.1007/s40857-019-00152-7

[28] C. Urbanietz and G. Enzner, "Binaural Rendering of Dynamic Head and Sound Source Orientation Using High-Resolution HRTF and Retarded Time," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, apr 2018, pp. 566–570. [Online]. Available: https://ieeexplore.ieee.org/document/8461343/

[29] E. Thuillier, H. Gamper, and I. J. Tashev, "Spatial Audio Feature Discovery with Convolutional Neural Networks," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2018-April, pp. 6797–6801, 2018.

[30] F. Keyrouz, "A novel robotic sound localization and separation using non-causal filtering and Bayesian fusion," *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, vol. 2016-Novem, pp. 0–5, 2016.

[31] P. Dadvar and M. Geravanchizadeh, "Robust binaural speech separation in adverse conditions based on deep neural network with modified spatial features and training target," *Speech Communication*, vol. 108, no. December 2018, pp. 41–52, 2019. [Online]. Available: https://doi.org/10.1016/j.specom.2019.02.001

[32] G. Kestler, S. Yadegari, and D. Nahamoo, "Head Related Impulse Response Interpolation and Extrapolation Using Deep Belief Networks," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2019-May, pp. 266–270, 2019.

[33] M. Tamulionis and A. Serackis, "Comparison of Multi-Layer Perceptron and Cascade Feed-Forward Neural Network for Head-Related Transfer Function Interpolation," *2019 Open Conference of Electrical, Electronic and Information Sciences, eStream 2019 - Proceedings*, pp. 1–4, 2019.

[34] Z. Haraszy, S. Micut, V. Tiponut, and T. Slavici, "Multi-subject head related transfer function generation using artificial neural networks," *International Conference on Systems - Proceedings*, vol. 1, no. Volume II, pp. 399–404, 2010.

[35] H. Hu, L. Zhou, H. Ma, and Z. Wu, "HRTF personalization based on artificial neural network in individual virtual auditory space," *APPLIED ACOUSTICS*, vol. 69, no. 2, pp. 163–172, feb 2008.

[36] W. Tu, Y. Yang, B. Du, J. Zheng, and S. Zhai, "Towards a real-time production of immersive spatial audio of high individuality with an RBF neural network," *Journal of Parallel and Distributed Computing*, vol. 131, pp. 120–129, 2019.

[37] C. Pang, H. Liu, and X. Li, "Multitask Learning of Time-Frequency CNN for Sound Source Localization," *IEEE Access*, vol. 7, pp. 40 725–40 737, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8668414/

[38] F. C. Tommasini, O. A. Ramos, M. X. Hüg, and F. Bermejo, "Usage of spectral distortion for objective evaluation of personalized HRTF in the median plane," *International Journal of Acoustics and Vibrations*, vol. 20, no. 2, pp. 81–89, 2015.

[39] S. K. Zieliński and H. Lee, "Automatic spatial audio scene classification in binaural recordings of music," *Applied Sciences (Switzerland)*, vol. 9, no. 9, 2019.

[40] L. Ghamdan, M. A. Ismail Shoman, R. A. Elwahab, and N. A. El-Hadid Ghamry, "Position estimation of binaural sound source in reverberant environments," *Egyptian Informatics Journal*, vol. 18, no. 2, pp. 87–93, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.eij.2016.05.002

[41] R. Urraca, E. Martinez-de Pison, A. Sanz-Garcia, J. Antonanzas, and F. Antonanzas-Torres, "Estimation methods for global solar radiation: Case study evaluation of five different approaches in central Spain," *Renewable and Sustainable Energy Reviews*, vol. 77, no. March, pp. 1098–1113, 2017.

[42] Y. X. Tan, J. L. Ng, and Y. F. Huang, "Estimation of missing daily rainfall during monsoon seasons for tropical region: A comparison between ann and conventional methods," *Carpathian Journal of Earth and Environmental Sciences*, vol. 15, no. 1, pp. 103–112, 2020.

[43] J. Ma, Y. Ding, J. C. Cheng, F. Jiang, V. J. Gan, and Z. Xu, "A Lag-FLSTM deep learning network based on Bayesian Optimization for multi-sequential-variant PM2.5 prediction," *Sustainable Cities and Society*, vol. 60, no. November 2019, p. 102237, 2020. [Online]. Available: https://doi.org/10.1016/j.scs.2020.102237