Audio and Communications Signal Processing Group (GTAC)

HEAD OF THE GROUP RESEARCH REPORT

The Audio and Communications Signal Processing Group (known by their acronym GTAC from its Spanish name Grupo de Tratamiento de señal en Audio y Comunicaciones) has developed its research during the scholar year 2021-22 mainly on active sound field control, personal sound zones, spatial audio perception and rendering, and sound quality improvement for multi-channel audio systems. GTAC has carried out several research projects and has published their most relevant results in several scientific journals and conference proceedings. Two new national projects have begun this year, "Adaptive Sound-processing Technologies foR soundfIeld Deployment: algorithms, tools and test beds (ASTRID)." and "Characterization of dynamic acoustic environments using machine learning (DYNAMIC)". Whereas, the national project "Intelligent Spatial Audio Synthesis and Customization (ISLA-THESON)" and the regional project "Smart Social Computing and Communication (COMTACTS)" are in halfway through their completion. On the other hand, the national projects "Dynamic Acoustic Networks for Changing Environments (DANCE)" has ended with great success, achieving its main objectives by creating several demonstrators for indoor dynamic real-life scenes. For instance, personal audio system for indoor environments and massive multichannel noise reduction in open-plan offices. More details of the projects' achievements are shown at the "Ongoing Projects" section.

Figure 1. Listening room overview.

With regards to the GTAC audio facilities, a large listening room of 40 m2 is totally equipped with top-notch audiovisual and control instrumentation (see Fig. 1).

Moreover, car seats surrounded by sound transducers are placed in this room to measure both objectively and perceptually, local sound zones around listener heads in enclosures (such as a cabin of a public transport or a living room, Fig. 2). As singular equipment, a self-designed robotic X-Y-Azimuth platform (see Fig. 3), which can support and move any sound recording or emitting device, is available, i.e., arrays of microphones, loudspeaker or audio-head. This platform will be used to characterize and monitor dynamic acoustic zones in indoor environments, as well as experiment with moving sound sources.

Figure 2. Car seats with transducers.

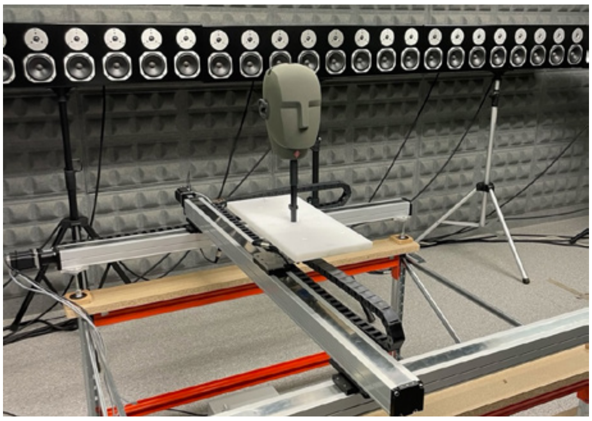

Figure 3. Robotic X-Y-Azimuth platform.

On the other hand, the laboratory for perceptual spatial sound of Fig. 4 allows measuring Head-Related Transfer Functions (HRTF) of any person with very high precision, in such a way that spatial sound can be rendered to a that particular person with high fidelity. The HRTF is somehow a personal acoustic fingerprint that changes from one person to another. By using individualized HRTFs, we can generate a virtual sound that is indistinguishable from reality. The loudspeaker array is formed by a 4-meter-diameter circular array of 72 loudspeakers placed in the same horizontal plane, plus two sets of 8 loudspeakers, one placed in the ceiling and one on the floor.

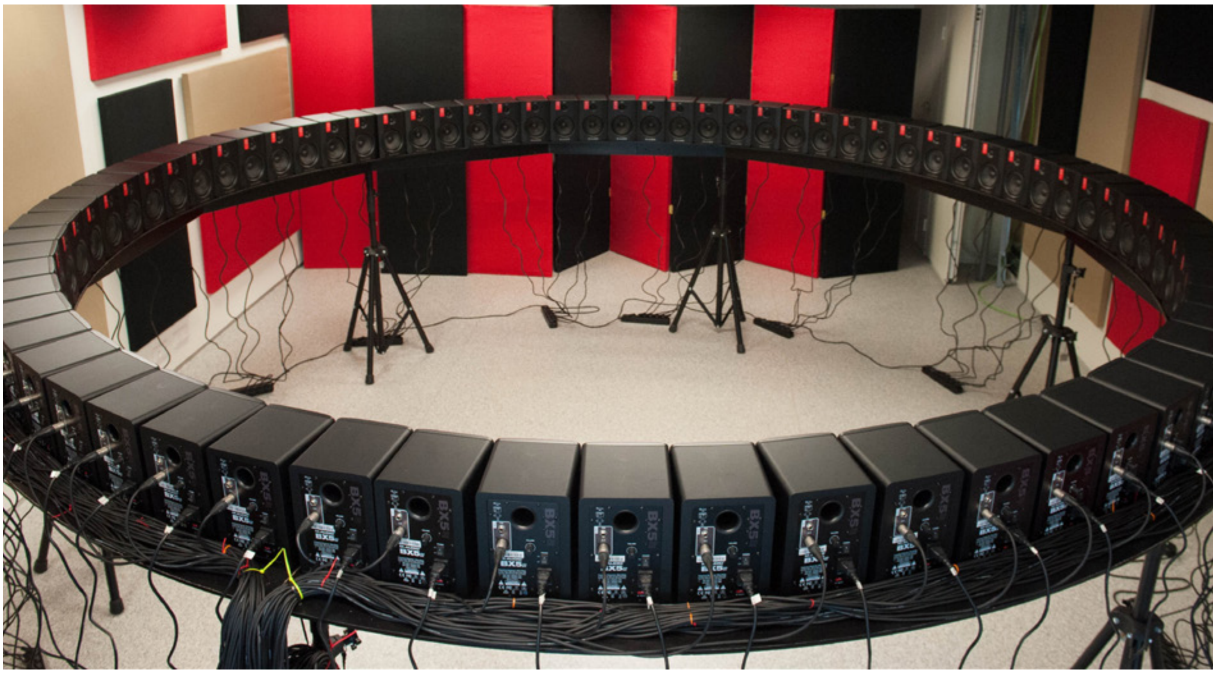

Figure 4. General view of the perceptual spatial sound laboratory.

1.- Project activities

In the following we describe the main ongoing projects that are being developed by GTAC researchers.

***Title*: INTELLIGENT SPATIAL AUDIO: SYNTHESIS AND CUSTOMIZATION (ISLA-THESON)**

Funded by: Spanish Ministry of Science, Innovation and University. 2019-2022.

The sound industry has been experiencing profound changes in recent years under the perspective of three complementary approaches: the individual, the group and the contents. Due the advances in virtual reality, mobile devices, video games and innmersive 3D movies, the spatial audio is today a discipline that attracts the attention of the industry. In this context, spatial audio systems try to accurately recreate the acoustic sensations that a listener would perceive within a real listening environment. Moreover, the use of headphones has spread enormously, and the need to reproduce highly realistic spatial sound through them is a great opportunity for the industry. For a very immersive experience, the sound must be customized for each individual based on their anatomy, in particular the head and pinna shape, which define their particular Head-Related Transfer Function (HRTF). Measuring a subject's HRTF is still a costly process that requires specialized facilities and finding an indirect way to get individualized HRTF is required. At ITEAM, we have built a new facility to measure HRTFs of real subjects in an efficient way (Fig.4).

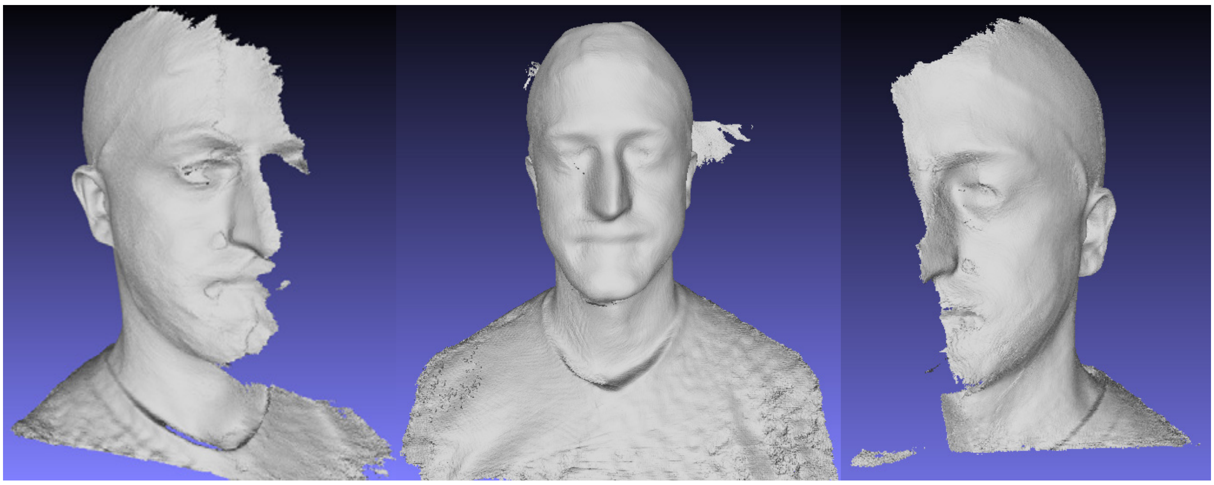

Fig. 5 Three different scans of the subject.

By employing Deep Learning techniques and photographs of the ear/head, we have achieved an HRTF personalization of better quality than previous methods. Previously, a new system has been constructed for the capture and extraction of individual anthropometric parameters from photographs. To this end, work is being done on the creation of 3D models through mobile devices that are equipped with depth cameras. The results obtained by combining both objective measurements (individual HRTF and anthropometric parameters) with deep learning techniques, can be evaluated by means of subjective perceptual tests. By using an individualized HRTF, we can generate a virtual sound indistinguishable from reality. This will in turn allow mobile devices to incorporate personalized responses for their direct application in 3D sound, virtual and augmented reality, video games, etc.

A smartphone with a 3D scan application was used to obtain these 3D models. The smartphone had a dot projector which could project more than 30000 infrared dots on the subject and a TrueDepth camera which could capture the images for the 3D reconstruction. Then, the 3D scan application could convert these images into a 3D model with a resolution range of 0.5 mm to 8.0 mm. Three different scans were taken, one of the head and two of both ears (see Fig. 5). Finally, in post-processing unwanted parts of the model were removed, holes were filled, and the scans were merged (see Fig.6).

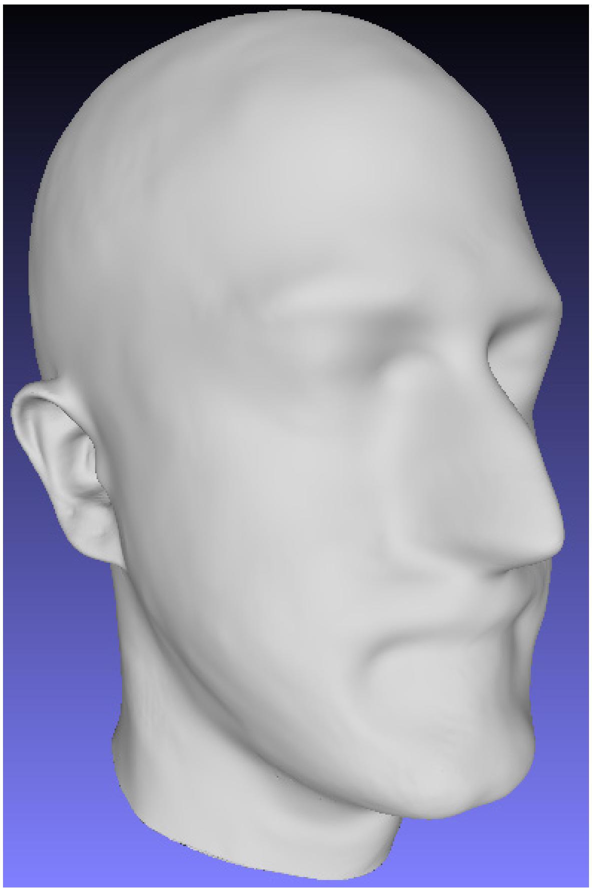

Figure. 6 Final model of the subject.

***Title*: SMART SOCIAL COMPUTING AND COMMUNICATION (in Spanish: COMUNICACIÓN Y COMPUTACIÓN INTELIGENTES Y SOCIALES - COMTACTS)**

Webpage: www.comtacts.upv.es

Funded by: Prometeo Call. Regional Government – Generalitat Valenciana. 2019-2023.

The advances made in the field of distributed computing and the hardware-software available right now make possible to develop powerful systems to process and exchange information, and at the same time, able to interact with the environment through numerous sets of transducers. These transducers, in turn, provide an ever-increasing volume of signals and data, making possible a more precise knowledge of the social and physical environment of the human beings' daily life.

On the other hand, let us consider the boom in applications arising from computing and communication devices for personal use, and their massive use with the advance of communications; some highlighted applications are human-machine interaction, control systems, location and tracking systems, telepresence, automatic classification, high-speed communications, diagnostic assistance systems, etc. Within this framework, intelligent and social computing and communication is defined as the hybrid mix of the two disciplines in order to face challenges of high socio-economic interest. Science is used for the purpose of communications and computing, but taking into account ubiquity, versatility, scalability, efficiency and cooperative processing of heterogeneous computing and data acquisition device networks.

COMTACTS project considers the physical aspects of computing, signal processing, energy consumption, technology, communication, etc., particularly in distributed, collaborative scenarios where massive and heterogeneous data are provided. In this way, COMTACTS addresses the design, development and implementation of products, systems, programs and algorithms for signal processing and communications, which make use of state-of-the-art architectures, advanced computing and efficient communications within the framework of intelligent computing and communication aimed at tackling social challenges.

***Title*: DYNAMIC ACOUSTIC NETWORKS FOR CHANGING ENVIRONMENTS (DANCE)**

Webpage: www.dance.upv.es

Funded by: Spanish Ministry of Science, Innovation and University. 2019-2022.

DANCE is a coordinated project that has developed distributed algorithms and systems to deal with different audio applications under the common frame of dynamic scenarios. Some of their tasks are: self-localization of nodes' positions, estimation of dynamic room impulse responses (RIRs) and inverse filters, fast adaptation and/or implementation over a distributed and heterogeneous network, characterization and control strategies adapted to the environments where control or listening points may vary with time, development of multiuser perceptual equalization methods to improve the listening experience in presence of undesired ambient noises. Additionally, emerging computing tools have been used to meet the real-time requirements of audio rendering and control in time-varying scenarios.

The DANCE project has included the development of four testbeds. The first one allows the design of personal sound zones (PSZ). The aim is to render a target soundfield in the "bright" zone while having control over the mean acoustic energy in the "dark" (quiet) zone. Examples are: watching TV and simultaneously listening to different languages in different positions, improving the listening experience in any room, tracking the listener over the home. The second demonstrator consists of a network of acoustic nodes that work together and simultaneously for the classification of sounds in the city using a Raspberry as a computing device. The third testbed comprises a massive multichannel noise reduction system for open-plan offices. The goal is to reduce the annoyance caused by the ambient noise and speech produced by other workers in open working spaces through their masking with pleasant sounds. The four testbed includes the development of perceptual equalization methods to improve the listening experience in presence of undesired broadband noises with multiple listeners.

***Title:* CHARACTERIZATION OF DYNAMIC ACOUSTIC ENVIRONMENTS USING MACHINE LEARNING (DYNAMIC)**

Funded entity and duration: Spanish Ministry of Science, Innovation and University. 2022-2025.

The overall user experience of media and entertainment is expanding in terms of the types of services, the environment where it is consumed, and the devices used to access them. However, the COVID-19 pandemic has abruptly changed our habits and most of the audiovisual content is consumed at homes instead of at theaters. Moreover, innovative entertainment experiences such as streaming music concerts have joined the traditional video (films and user-generated content) streaming and gaming services. At this point, it is necessary to develop intelligent sound reproduction systems that can provide in a natural way (without headphones) good-quality sound for a full immersive experience. The primary goal of the DYNAMIC project is to investigate on sound space control applications for real and dynamic environments using novel machine and deep learning techniques, aiming at maximum performance and feasibility.

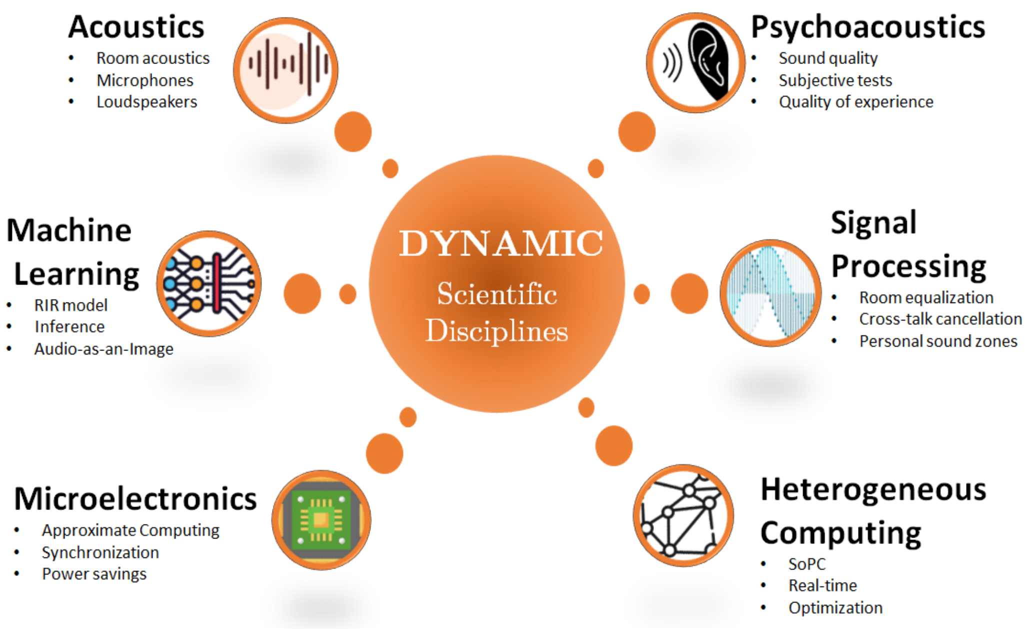

Figure 7. Multidisciplinary diagram of the coordinated DYNAMIC project.

The acoustic environment is the physical medium by which the sound propagates, and its physical characteristics have been studied for decades. Thanks to the digital approach to the physical modelling of sound propagation, the acoustic environment can be described by its room impulse response (RIR). To deal with dynamic acoustic environments where each time a listener moves a new RIR is generated, all the possible RIRs of the room should be known a priori. In this project, a new approach to the digital characterization of dynamic acoustic environments is proposed based on machine learning (ML) and, especially, deep Learning (DL) algorithms. DL has successfully addressed a wide range of applications in the audio field, as source separation, speech enhancement and acoustic echo cancellation. Most of the ML models that try to solve these problems have one thing in common: the input to the neural network is an image that represents the audio content, usually the short-term Fourier Transform (STFT). This will be the main approach to tackle the problem.

Title: **ADAPTIVE SOUND-PROCESSING TECHNOLOGIES FOR SOUNDFIELD DEPLOYMENT: ALGORITHMS, TOOLS AND TEST BEDS (ASTRID).**

Webpage of the project: https://gtac-iteam-upv.github.io/Astrid/

Funding entity and duration: Spanish Ministry of Science, Innovation and University. 2022-2025.

ASTRID is proposed as a project that involves several PhD researchers in the field of signal processing and computing, with long-term, well-established collaboration between them and showed experience to confidently address a suitably selected range of scientific and technological challenges in the field of signal processing of sound signals and computing. Furthermore, this project exhibits a strong vocation for research excellence, internationalization, training and transfer. ASTRID Project promotes multidisciplinary research since it mobilizes complementary knowledge from various scientific fields, mainly: signal processing, telematics, computing, applied matematics and applied physics; to get its objectives, which are oriented to: research, dissemination, formation and transfer; towards finding solutions to society's problems, through publication of research results in forums with high scientific and technological impact, technology transfer and internationalization of activities.

The general objective of ASTRID is to contribute to sound-field deployment systems by the design and development of test beds, algorithms and computational kernels, that improve the performance, increase the resiliency, and reduce the energy consumption in order to transfer knowledge and tools to the productive sector. ASTRID will address complex sound deployment scenarios along three research fields: Fast multichannel adaptive algorithms, Distributed and collaborative systems and Psychoacoustic aspects of listening; focused on four related application target domains: Active sound field control, Personal sound zones and spaces, and Computational and mathematical tools for sound processing.

That can be settled down with the following ones: Improve the sound deployment systems in complex scenarios when listener or reproduction zones moves or change. Improve the sound deployment systems in distributed scenarios. Improve human perception of sound deployment systems. Develop and implement a Road Active Noise Control system.

The demand for a quieter and healthier human environment is growing rapidly and needs experts possessing a comprehensive knowledge and practical experience in sound and noise control technology. Currently, market-available solutions are not sufficient in many cases. Regulatory guidelines also favour demand in the noise control system market. But through the completion of ASTRID new feasible and comprehensive noise reducing systems will be developed and delivered, for both indoor environments and vehicle/aircraft cabins, boosting the advancement of European companies in this field and allowing them to take advantage of the growing noise control market.

2.- Research results

The most important results of the GTAC publications over the past year are summarized in the following. For a more detailed description, visit our webpage: www.gtac.upv.es where a complete list of projects and papers can be found.

2.1.- Featured Journal Publications

- ***Transfer functions of FXLMS-based Multi-channel Multi-tone Active Noise Equalizers.*** *Miguel Ferrer, María de Diego, Gema Piñero, Amin Hassani, Marc Moonen, Alberto González, Electrical Engineering and Systems Science, Audio and Speech Processing, 2022. DOI: [10.48550/arXiv.2207.01102](https://doi.org/10.48550/arXiv.2207.01102).*

  Abstract: Multi-channel Multi-tone Active Noise Equalizers can achieve different user-selected noise spectrum profiles even at different space positions. They can apply a different equalization factor at each noise frequency component and each control point. Theoretically, the value of the transfer function at the frequencies where the noise signal has energy is determined by the equalizer configuration. In this work, we show how to calculate these transfer functions with a double aim: to verify that at the frequencies of interest the values imposed by the equalizer settings are obtained, and to characterize the behavior of these transfer functions in the rest of the spectrum, as well as to get clues to predict the convergence behaviour of the algorithm. The information provided thanks to these transfer functions serves as a practical alternative to the cumbersome statistical analysis of convergence, whose results are often of no practical use.

- ***Low-complexity soft ML detection for generalized spatial modulation.*** *M. Angeles Simarro, Víctor Manuel García Mollá, Francisco José Martínez Zaldívar, Alberto Gonzalez, Signal Processing, vol. 196, July 2022. DOI:* 10.1016/j.sigpro.2022.108509.

  Abstract: Generalized Spatial Modulation (GSM) is a recent Multiple-Input Multiple-Output (MIMO) scheme, which achieves high spectral and energy efficiencies. Specifically, soft-output detectors have a key role in achieving the highest coding gain when an error-correcting code (ECC) is used. Nowadays, soft-output Maximum Likelihood (ML) detection in MIMO-GSM systems leads to a computational complexity that is unfeasible for real applications; however, it is important to develop low-complexity decoding algorithms that provide a reasonable computational simulation time in order to make a performance benchmark available in MIMO-GSM systems. This paper presents three algorithms that achieve ML performance. In the first algorithm, different strategies are implemented, such as a preprocessing sorting step in order to avoid an exhaustive search. In addition, clipping of the extrinsic log-likelihood ratios (LLRs) can be incorporating to this algorithm to give a lower cost version. The other two proposed algorithms can only be used with clipping and the results show a significant saving in computational cost. Furthermore clipping allows a wide-trade-off between performance and complexity by only adjusting the clipping parameter.

- ***Weighted pressure matching with windowed targets for personal sound zones.*** *Vicent Molés-Cases, Stephen J. Elliott, Jordan Cheer, Gema Piñero, Alberto Gonzalez, The Journal of the Acoustical Society of America, vol. 151, 2022. DOI:* 10.1121/10.0009275.
  Abstract: Personal sound zones (PSZ) systems use an array of loudspeakers to render independent audio signals to multiple listeners within a room. The performance of a PSZ system, designed using weighted pressure matching, depends on the selected target responses for the bright zone. In reverberant environments, the target responses are generally chosen to be the room impulse responses from one of the loudspeakers to the control points in the selected bright zone. This approach synthesizes the direct propagation component and all the reverberant components in the bright zone, while minimizing the energy in the dark zone. We present a theoretical analysis to show that high energy differences cannot be achieved for the diffuse reverberant components in the bright and dark zones, and so trying to synthesize these components in the bright zone does not lead to the best performance. It is then shown that the performance can be improved by using windowed versions of these measured impulse responses as target signals, in order to control which reverberant components are synthesized in the bright zone and which are not. This observation is supported by experimental measurements in two scenarios with different levels of reverberation.

- ***Parallel signal detection for generalized spatial modulation MIMO systems.*** *Victor M. Garcia-Molla, M. Angeles Simarro, F. J. Martínez-Zaldívar, Murilo Boratto, Pedro Alonso, Alberto Gonzalez, The Journal of Supercomputing, vol. 78, pp. 7059 – 7077, 2022. DOI:* 10.1007/s11227-021-04163-y.

  Abstract: Generalized Spatial Modulation is a recently developed technique that is designed to enhance the efficiency of transmissions in MIMO Systems. However, the procedure for correctly retrieving the sent signal at the receiving end is quite demanding. Specifically, the computation of the maximum likelihood solution is computationally very expensive. In this paper, we propose a parallel method for the computation of the maximum likelihood solution using the parallel computing library OpenMP. The proposed parallel algorithm computes the maximum likelihood solution faster than the sequential version, and substantially reduces the worst-case computing times.

- ***Compensating first reflections in non-anechoic head-related transfer function measurements.*** *Jose J. Lopez,* *Pablo Gutierrez-Parera, Máximo Cobos. Applied Acoustics, vol. 188, 108523, 2022. DOI: [10.1016/j.apacoust.2021.108523](https://doi.org/10.1016/j.apacoust.2021.108523).*

Abstract: Personalized Head-Related Transfer Functions (HRTFs) are needed as part of the binaural sound individualization process in order to provide a high-quality immersive experience for a specific user. Signal processing methods for performing HRTF measurements in non-anechoic conditions are of high interest to avoid the complex and inconvenient access to anechoic facilities. Non-anechoic HRTF measurements capture the effect of room reflections, which should be correctly identified and eliminated to obtain HRTFs estimates comparable to ones acquired in an anechoic setup. This paper proposes a sub-band frequency-dependent processing method for reflection suppression in non-anechoic HRTF signals. Array processing techniques based on Plane Wave Decomposition (PWD) are adopted as an essential part of the solution for low frequency ranges, whereas the higher frequencies are easily handled by means of time-crop windowing methods. The formulation of the model, extraction of parameters and evaluation of the method are described in detail. In addition, a validation case study is presented showing the suppression of reflections from an HRTF measured in a real system. The results confirm that the method allows to obtain processed HRTFs comparable to those acquired in anechoic conditions.

- ***Interaural time difference individualization in HRTF by scaling through anthropometric parameters.*** *Pablo Gutierrez-Parera, Jose J. Lopez, Javier M. Mora-Merchan, Diego F. Larios. EURASIP Journal on Audio, Speech, and Music Processing, Article number: 9, 2022. DOI: [10.1186/s13636-022-00241-y](https://doi.org/10.1186/s13636-022-00241-y).*

Abstract: Head-related transfer function (HRTF) individualization can improve the perception of binaural sound. The interaural time difference (ITD) of the HRTF is a relevant cue for sound localization, especially in azimuth. Therefore, individualization of the ITD is likely to result in better sound spatial localization. A study of ITD has been conducted from a perceptual point of view using data from individual HRTF measurements and subjective perceptual tests. Two anthropometric dimensions have been demonstrated in relation to the ITD, predicting the subjective behavior of various subjects in a perceptual test. With this information, a method is proposed to individualize the ITD of a generic HRTF set by adapting it with a scale factor, which is obtained by a linear regression formula dependent on the two previous anthropometric dimensions. The method has been validated with both objective measures and another perceptual test. In addition, practical regression formula coefficients are provided for fitting the ITD of the generic HRTFs of the widely used Brüel & Kjær 4100 and Neumann KU100 binaural dummy heads.

2.2.- Featured Conference Proceedings

- ***Perceptual active noise equalization with virtual microphones.*** *Juan Estreder, Miguel Ferrer, Maria de Diego, Gema Piñero, Alberto Gonzalez, 28th International Congress on Sound and Vibration (ICSV28), Singapur, July 2022.*

- **Conditional Generative Adversarial Networks for Acoustic Echo Cancellation**, F. Pastor-Naranjo, R. del Amor, J. Silva-Rodríguez, M. Ferrer, G. Piñero, V. Naranjo (2022). *30th European Signal Processing Conference, EUSIPCO 2022.*

  https://eurasip.org/Proceedings/Eusipco/Eusipco2022/pdfs/0000085.pdf

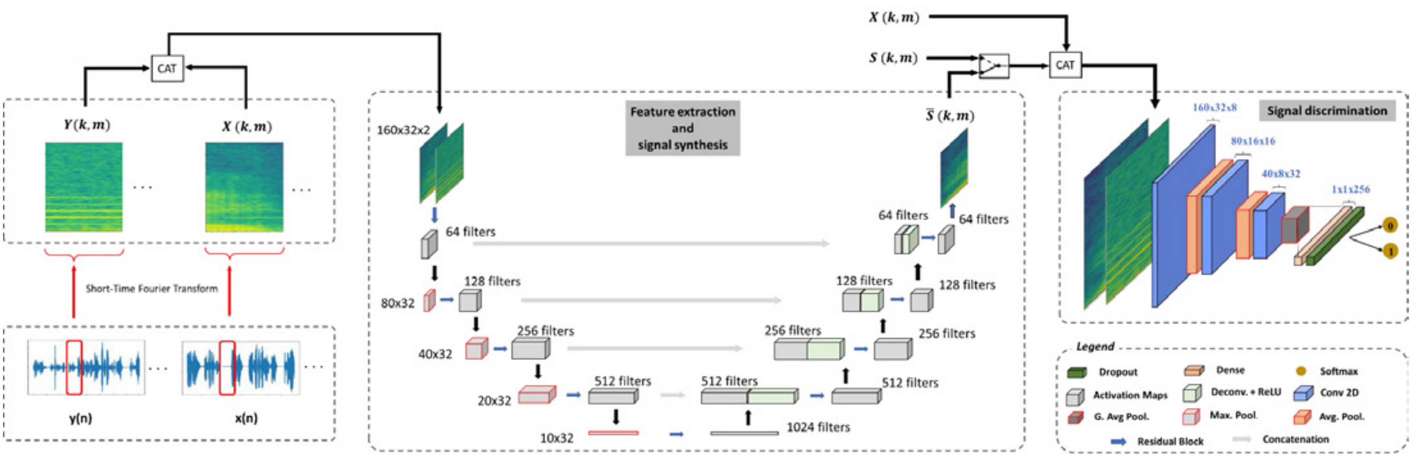

Figure 8. Proposed framework to perform acoustic echo cancellation.