AUDIO AND COMMUNICATIONS SIGNAL PROCESSING GROUP

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 2019-20 mainly on active noise control, 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. In particular, the Cátedra Telefonica-UPV project "Sound-Aided Smart Environments for the City, Home and Nature (SSEnCe)" is reaching its final stage with great success, achieving their objectives by creating a demonstrator that allows detecting and classifying acoustic events, for home environments, Smart Cities and Natural parks. On the other hand, the national projects "Dynamic Acoustic Networks for Changing Environments (DANCE)" and "Intelligent Spatial Audio Synthesis and Customization (ISLA-THESON)" are in halfway through their completion. More details of their achievements are shown at the "Ongoing Projects" section, but in the following, we want to highlight the most novel developments of the GTAC during this last year.

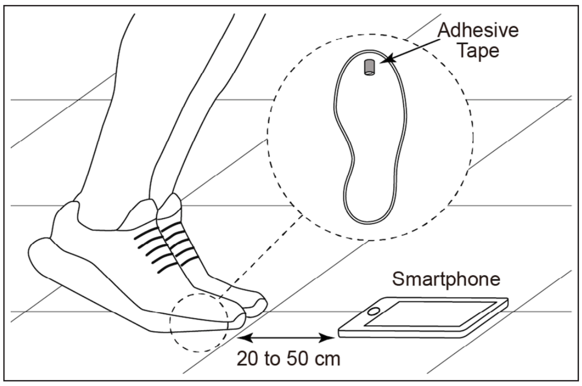

Figure 1. Experimental setup of jump recordings with the microphone of a smartphone.

Figure 2. Subjective test on 3D sound perception using individualized HRTFs.

A new audio application in Sports (see Fig. 1) has been developed with the collaboration of researchers from the University of Alicante. This is an audio-based system acting as a trustworthy instrument to accurately measure the height of a jump. It can be easily automated as a mobile app to facilitate its use both in laboratories and in the field.

With regards to the GTAC facilities, the new laboratory for perceptual spatial sound is finishing their equipment. It 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 (see Fig.2). The HRTF is in 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. As it can be seen from Fig. 2, 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.

1.- PROJECT ACTIVITIES

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

ONGOING PROJECTS

Title: Sound-Aided Smart environments for the city, home and nature (SSEnCe)

Webpage of the project: www.sound-aided-IOT.webs.upv.es

Funding entity and duration: Cátedra Telefónica UPV. 2017-2020

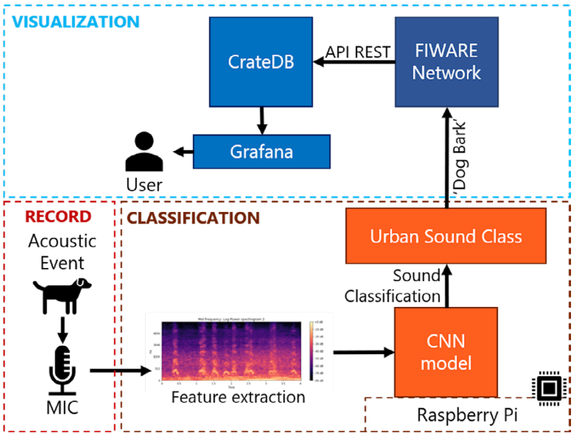

Figure 3. Scheme of the WASN used for cities' sound classification.

The SSEnCe project aims to encourage the development and dissemination of real and practical prototypes focused on the concept of intelligence for the Internet of Things (IoT). Particularly, the project will develop applications mainly addressed to obtain acoustic information of the environment. A second main objective of this project is the creation of an observatory of technological demonstrators developed by national and international research groups related to the acoustic-aided IoT.

We have developed within the frame of the project a demonstrator of an environmental sound classifcafier (ESC) of city sounds based on a wireless acoustic sensor network (WASN) whose nodes are low-cost devices, whose scheme is shown in Fig. 3. The WASN recognizes a set of sound events or classes from urban environments. Their nodes are formed by Raspberry Pi devices equipped with outdoor microphones, and they not only record the ambient sound, but can also process and recognize a sound event by means of deep learning (convolutional neural network (CNN) model in Fig. 3). To our knowledge, this is the first WASN running a CNN classifier at their own nodes and not using cloud or edge computing to get the record signals classified. In our WASN, the nodes send the resulting probability of every sound class to the server, so the data can be displayed in a map. Such WASNs have many advantages as monitoring system: they are cheap compared to other monitoring systems, they can be easily deployed and they can work day and night. An additional advantage of our WASN is that uses the open standard FIWARE in their communication network, so the whole system can be replicated without the need of proprietary software or hardware. Fig. 4 shows the system dashboard where the data are visualized, including the location of the nodes over the map of Valencia city.

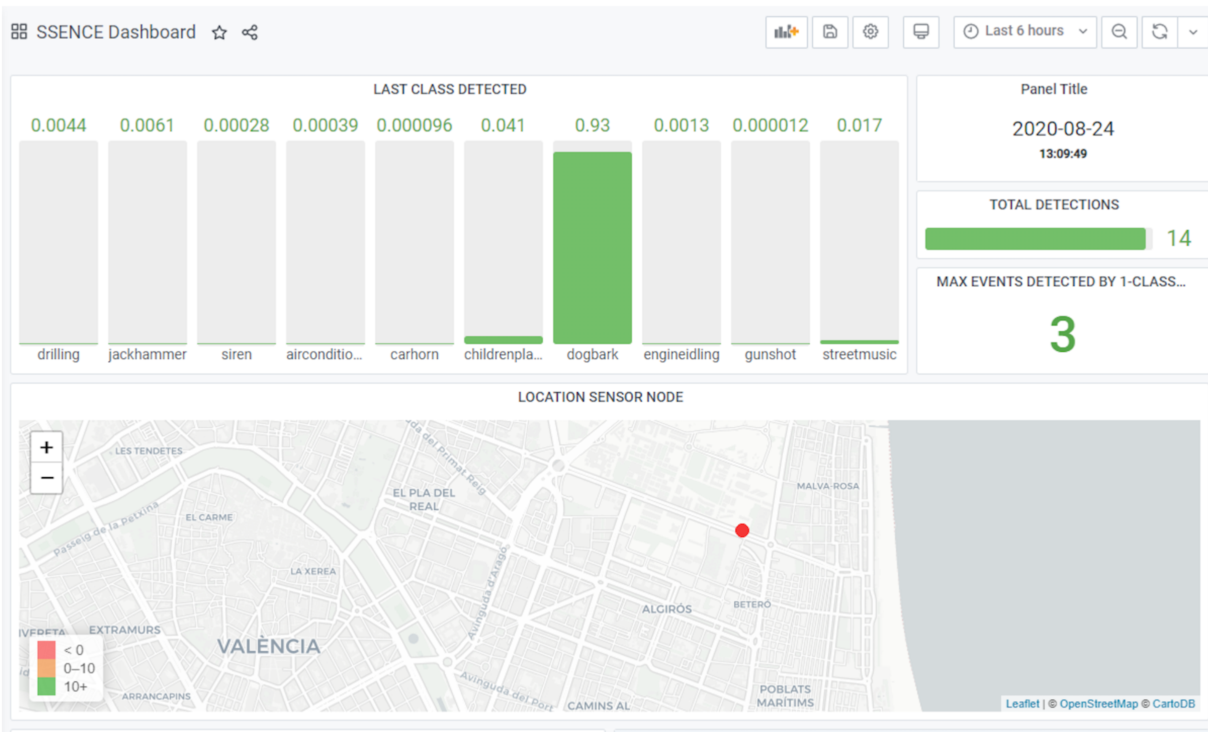

Figure 4. Dashboard of the WASN monitoring system in Valencia city.

Title: Dynamic acoustic networks for changing environments (DANCE)

Webpage: www.dance.upv.es

Funded by: Spanish Ministry of Science, Innovation and University. 2019-2021.

DANCE is a coordinated project that will develop distributed algorithms and systems to deal with different audio applications under the common frame of dynamic scenarios. Some of their tasks will be: self-localization of nodes' positions, estimation of dynamic room impulse responses (RIRs) and inverse filters, characterization and control of time-varying acoustic wave fields, etc. Additionally, emerging computing tools will be used to meet the real-time requirements of audio rendering and control in time-varying scenarios.

The DANCE project includes the development of two testbeds in the GTAC audio laboratory: (1) The first one employs sub-band filtering and optimized filter bank computation in the time domain for 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. The PSZ system presents a great versatility since it can be adapted and optimized according to the frequency content of the audio signals, the characteristics of the room and the typology and location of the transducers used. (2) The second testbed consists in a massive multichannel noise reduction for open-plan offices. The aim is to reduce the annoyance caused by the ambient noise and speech in open working spaces through their masking with pleasant sounds. These pleasant sounds (waterfall, birdsongs, pink noise,...) will make the annoying ambient sounds almost inaudible thanks to the masking properties of the human hearing system.

For this purpose, we have designed a robotic X-Y-Azimuth platform (see Fig. 5) able to support and move any recording or audio emitting device, i.e. arrays of microphones or audio-heads as shown in Fig. 5. The platform is controlled by its own software, but with it can also be controlled from Matlab. It can operate within an area of 1.3x1.3m (X-Y) and supports a whole azimuth range of 360°. This platform will be used to characterize dynamic acoustic zones in indoor environments.

Title: Intelligent spatial audio: synthesis and customization (ISLA-THESON)

Funded by: Spanish Ministry of Science, Innovation and University. 2019-2021.

The sound industry has been experiencing profound changes in recent years under the

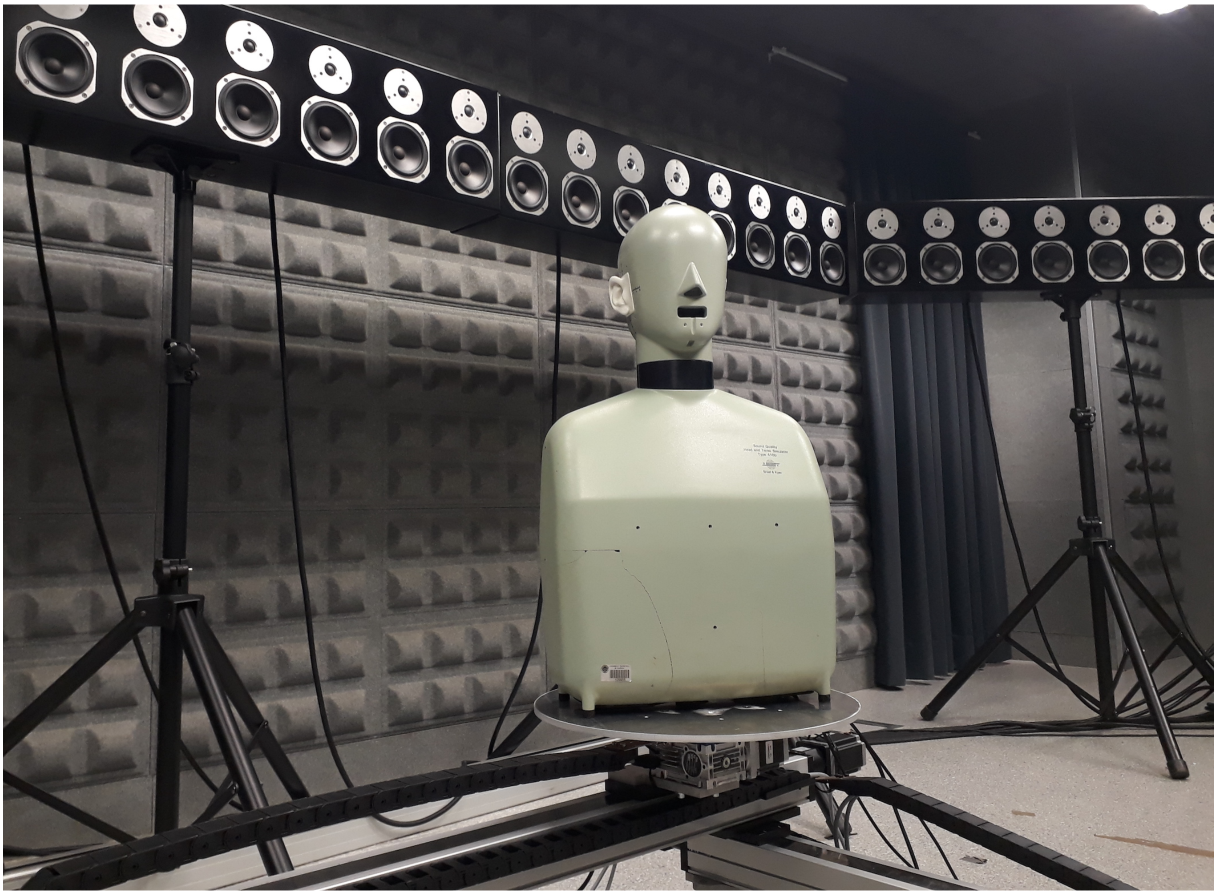

Figure 5. Robotic platform for the measurement of dynamic scenarios.

Figure 6. Objective measurements of the HRTF.

perspective of three complementary approaches: the individual, the group and the contents. 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 its indirect estimation remains an unresolved problem. At ITEAM, we have built a new facility to measure HRTFs of real subjects in an efficient way (Fig. 6). By employing Deep Learning techniques and photographs of the ear/head, we have achieved an HRTF personalization of better quality than previous methods. To this end, a new system has been constructed for the capture and extraction of individual anthropometric parameters from photographs (Fig. 7). 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 as was shown in Fig. 2. 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.

On the other hand, the sound and entertainment industry has been redirected during the recent years to big live shows, where the spatialization of sound is still a challenge and an opportunity for using sound field synthesis algorithms to recreate virtual spaces. Array processing

Figure 7. Setup of the multi-camera system.

techniques should be developed to control the sound in different listening areas while synthesizing the different live sound objects (musicians, actors, presenters, effects, etc.), adapting the synthesis of each object to its own movement and achieving greater realism over the audience. Other scenarios such as museums, exhibitions, restaurants or smart homes would also benefit from the creation of independent audio zones, using similar techniques employing loudspeaker and sensor arrays.

Finally, from the contents point of view, this subproject will work on creating new methods for the analysis of audio and music based on Machine Learning, with application to synchronized audiovisual effects and live enriched events. The aim is to develop Machine Learning algorithms able to extract features from music and enable the synchronization of 3D animations, lights or lasers with the music.

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.

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

Personal Sound Zones by Subband Filtering and Time Domain. *Vicent Moles, Gema Piñero, Maria de Diego, Alberto Gonzalez, IEEE/ACM Trans. on Audio, Speech and Language Processing, vol. 28, pp. 2684 - 2696, 2020. DOI: 10.1109/TASLP.2020.3023628.*

Abstract: Personal Sound Zones (PSZ) systems aim to render independent sound signals to multiple listeners within a room by using arrays of loudspeakers. One of the algorithms used to provide PSZ is Weighted Pressure Matching (wPM), which computes the filters required to render a desired response in the listening zones while reducing the acoustic energy arriving to the quiet zones. This algorithm can be formulated in time and frequency domains. In general, the time-domain formulation (wPM-TD) can obtain good performance with shorter filters and delays than the frequency-domain formulation (wPM-FD). However, wPM-TD requires higher computation for obtaining the optimal filters. In this article, we propose a novel approach to the wPM algorithm named Weighted Pressure Matching with Subband Decomposition (wPM-SD), which formulates an independent time-domain optimization problem for each of the subbands of a Generalized Discrete Fourier Transform (GDFT) filter bank. Solving the optimization independently for each subband has two main advantages: 1) lower computational complexity than wPM-TD to compute the optimal filters; 2) higher versatility than the classic wPM algorithms, as it allows different configurations (sets of loudspeakers, filter lengths, etc.) in each subband.

Effects and applications of spatial acuity in advanced spatial audio reproduction systems with loudspeakers. *José Javier López Monfort, Pablo Gutiérrez-Parera, Lauri Savioja Applied Acoustics, vol. 161, 107179, 2020. DOI: 10.1016/j.apacoust.2019.107179*

Abstract: Spatial audio reproduction systems using loudspeakers produce coloration effects at high frequencies due to spatial interference between loudspeakers, in both those based on panning and those based on field synthesis. As a response to this problem and in order to reduce coloration, this paper studies the feasibility of an alternative approach where high frequencies are reproduced from a single loudspeaker, with a different direction from that of its panned low-frequency counterpart. Listening tests are conducted to investigate the localization and quality of the source in the case that frequencies higher than 1.5 kHz are reproduced from a different direction than the low frequencies. In this context, the human ability to discriminate the spatial direction of low/high frequency bands and the error in the perceived direction of arrival for different separation angles is evaluated and quantified. The resulting data has been analyzed with ANOVA, providing significant results that allow us to establish a threshold in the angular separation of the high and low frequency parts where subjects do not perceive source location artifacts. The term just noticeable band splitting angle (JNBSA) is defined and introduced. It represents the minimum angle of separation between high and low frequencies from which the listener starts to perceive artifacts in the reproduction of a sound source using loudspeakers.

Audio-Based System for Automatic Measurement of Jump Height in Sports Science. *Basilio Pueo, José Javier López Monfort, Jose M. Jimenez-Olmedo, Sensors, vol. 11, nª 19, pp. 2543- 2556, 2019. DOI: 10.3390/s19112543*

Abstract: Jump height tests are employed to measure the lower-limb muscle power of athletic and non-athletic populations. The most popular instruments for this purpose are jump mats and, more recently, smartphone apps, which compute jump height through manual annotation of video recordings to extract flight time. This study developed a non-invasive instrument that automatically extracts take-off and landing events from audio recordings of jump executions. An audio signal processing algorithm, specifically developed for this purpose, accurately detects and discriminates the landing and take-off events in real time and computes jump height accordingly. Its temporal resolution theoretically outperforms that of flight-time-based mats (typically 1000 Hz) and high-speed video rates from smartphones (typically 240 fps). A validation study was carried out by comparing 215 jump heights from 43 active athletes, measured simultaneously with the audio-based system and with of a validated, commercial jump mat. The audio-based system produced nearly identical jump heights than the criterion with low and proportional systematic bias and random errors. The developed audio-based system is a trustworthy instrument for accurately measuring jump height that can be readily automated as an app to facilitate its use both in laboratories and in the field.

On the use of many-core machines for the acceleration of a mesh truncation technique for FEM. *Jose A. Belloch, A. Amor-Martin, D. García-Donoro, Francisco J. Martinez, L. E. Garcia-Castillo, The Journal of Supercomputing, vol. 75, nª 3, pp. 1686-1696, 2019. DOI: 10.1007/s11227-018-02739-9.*

Abstract: Finite element method (FEM) has been used for years for radiation problems in the field of electromagnetism. To tackle problems of this kind, mesh truncation techniques are required, which may lead to the use of high computational resources. In fact, electrically large radiation problems can only be tackled using massively parallel computational resources. Different types of multi-core machines are commonly employed in diverse fields of science for accelerating a number of applications. However, properly managing their computational resources becomes a very challenging task. On the one hand, we present a hybrid message passing interface + OpenMP-based acceleration of a mesh truncation technique included in a FEM code for electromagnetism in a high-performance computing cluster equipped with 140 compute nodes. Results show that we obtain about 85% of the theoretical maximum speedup of the machine. On the other hand, a graphics processing unit has been used to accelerate one of the parts that presents high fine-grain parallelism.

2.2.- FEATURED CONFERENCE PROCEEDINGS

A Low-cost Wireless Acoustic Sensor Network for the Classification of Urban Sounds. *David Salvo, Gema Piñero, Pau Arce, Alberto Gonzalez, The Seventeenth ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor and Ubiquitous Networks, PEWASUN, 2020.*

Providing Spatial Control in Personal Sound Zones Using Graph Signal Processing. *Vicent Moles, Gema Piñero, Alberto Gonzalez, Maria de Diego 27th European Signal Processing Conference (EUSIPCO 2019), A Coruña, Spain, 2019.*

Towards low cost acoustic cameras for the Internet of Things. *José Javier López Monfort, Maximilian Becker, Carlos Alberto Hernandez Franco, 23rd International Congress on Acoustics (ICA 2019). (3329-3336). Aachen, Germany: German Acoustical Society.*