TECNIACÚSTICA
CUENCA 2023

EVALUATION OF A PERSONAL SOUND ZONE SYSTEM IMPLEMENTED WITH LOW-COST MULTIMEDIA DEVICES

Mathys Daniel [1], Daniel De La Prida [2], Laura Fuster [3], Gema Piñero [3], Luis A. Azpicueta-Ruiz [2]*

[1] ENSEA, Cergy, Francia
[2] Universidad Carlos III, Madrid, España
[3] Universitat Politècnica de València, Valencia, España

ABSTRACT

In a personal sound zone (PSZ) system, the subjective quality of the audio reaching the bright zone depends on the accuracy of the estimation of the acoustic response between each loudspeaker and the listener's position, known as the room impulse response (RIR). Typically, the estimation of the RIR must be carried out at a stage prior to the implementation of the PSZ system, using quasi-professional equipment. However, we use multimedia devices every day that could also perform RIR estimation, such as a smartphone connected to a wireless loudspeaker. This paper presents a comparison of a PSZ system for three sets of RIRs estimated by: 1) an Android device connected to a Bluetooth loudspeaker, 2) an array of Brüel&Kjaer (BK) microphones and the same loudspeaker, 3) an array of BK microphones and an array of JBL loudspeakers. The evaluation is performed using objective metrics and a subjective psychoacoustic test.

Keywords — Estimation of room impulse responses, Bluetooth loudspeakers, personal sound zones.

1. INTRODUCTION

The quest for personalized audio experiences within shared enclosures has paved the way for extensive research in acoustics, giving rise to the development of Personal Sound Zone (PSZ) systems [1]. This innovative application can create distinct and tailored audio regions within shared spaces, ensuring that each listener perceives sound according to their preferences, all without disturbing those around them. At the heart of the successful implementation of PSZ systems lies the critical process of accurately estimating the Room Impulse Response (RIR) between a set of loudspeakers and the positions of listeners.

RIR estimation for PSZ systems has predominantly relied upon the use of professional equipment. This approach ensures the capture of high-fidelity data, making it the gold standard for RIR estimation. However, the increasing ubiquity of multimedia devices in our daily lives has sparked an intriguing exploration: Can everyday gadgets and tools effectively estimate the RIR, at least for the purpose of building PSZs? Specifically, can a smartphone wirelessly connected to a Bluetooth loudspeaker be a feasible alternative to high-quality microphones and loudspeakers?

The goal of the work here presented is to compare the accuracy of RIR estimates achieved through these two (high-quality and low-cost) systems when they are used to design the filters required by the PSZ system. For this purpose, we will evaluate the mostly used Acoustic Contrast (AC) [2] to assess the quantitative performance of each set of RIR estimates. In parallel, we have carried out a psychoacoustic test where participants can evaluate their audio quality perception when the PSZ system is designed upon the RIR collected through each method.

The rest of the paper is as follows: Section 2 explains the methodology employed for data collection and analysis. We will present our findings in detail, offering valuable insights into the advantages and limitations of each RIR estimation method. Section 3 deals with the required preprocessing of the data to build a PSZ system. Section 4 presents the objective and subjective results, while Section 5 highlights the main conclusions of our study.

2. DATA COLLECTION

Three different collections or sets of impulse responses have been estimated for the same locations and with the same estimation method based on emitting a logarithmic chirp and

* ***Autor de contacto***: *gpinyero@iteam.upv.es*

This work has been partially funded by MCIN/AEI/ 10.13039/501100011033 and "ERDF A way of making Europe" through Grant PID2021-124280OB-C21.

estimating the RIR through a decorrelation process. This method has shown to have the least estimation error [3]. Fig.1 shows the setup of the PSZ application, where two zones of 21x21 cm have been measured with a distance of 7 cm between microphones. The loudspeaker array is formed by 8 loudspeakers separated 18 cm apart. In total, 256 (16x2x8) RIR measurements were conducted. Fig.1 also includes the distances of the loudspeaker and microphone arrays from the left wall of the room, and the distance between the first loudspeaker (from the left) and the left upper corner microphone of "Zone 1", whose RIR will be considered as the reference RIR in Section 3.

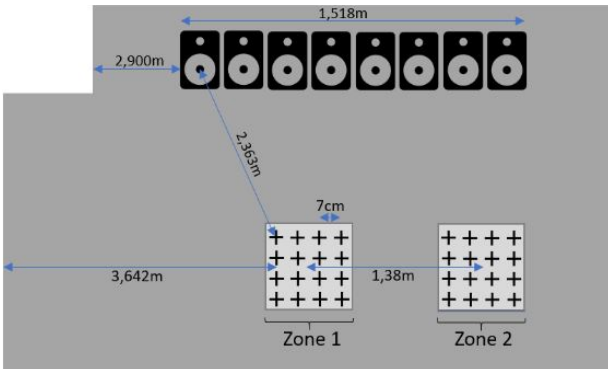


Figure 1. Data collection setup.

The distance between the respective central location of the two zones is 1,38 m. The whole room has dimensions 11,85x7,3x3 m and is mainly used as a laboratory. There are several tables, chairs, and other equipment all around the lab. Moreover, the wall at the south of the zones in Fig.1 is entirely made of crystal. The room presents an average reverberation time of T = 0,5 s. Three different combinations of devices and instrumentation has been used to estimate the RIRs, which will be explained in the following.

2.1. Android-based RIR Measurements

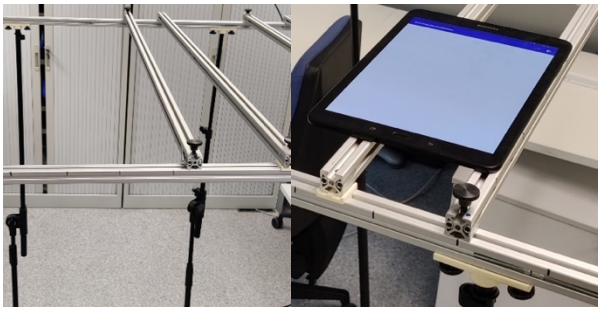
In this setup, the RIRs have been estimated using an Android tablet paired with a Bluetooth (BT) speaker. The tablet is a Samsung Galaxy Tab S3 and the BT loudspeaker model is a Yamaha NX-P100. The Bluetooth version is 2.1, which is quite old with respect to the latest 5.4 version [4], but as it will be explained in Section 3, BT random delays have been compensated and only the quality of the estimated RIRs will be considered in our study.

Fig.2 presents the Yamaha loudspeaker placed over two cardboard boxes at the 3rd position of the array. The boxes helped to place the loudspeaker membrane at a similar height as those of the professional array that will be described in Section 2.2.

Figure 2. Yamaha loudspeaker at the 3rd position of the professional array.

We have developed a proprietary app for Android devices able to perform the whole RIR estimation: It allows the user to connect to any available BT loudspeaker, and when it is ready, the tablet sends a chirp to the loudspeaker, records the sound emitted by the speaker with its built-in microphone, and obtains the RIR of the recorded signal through decorrelation.

Since there was only one tablet and one loudspeaker, we had to assure the accurate position of the tablet in the grid shown in Fig.1 for both zones. Therefore, we built a light structure of 1x1m (see Fig.3) that could be arranged to place the tablet in any point of the grid. These rails were marked at 7 cm intervals for both X and Y dimensions, such that the tablet microphone was located at the exact positions shown in Fig.1 for Zones 1 and 2.



(a) (b)

Figure 3. (a) Structure with holders and rails. (b) Tablet setting.

Summarizing, the steps carried out to obtain the RIRs have been: 1) the BT loudspeaker was placed in one of the positions of the professional array, as shown in Fig.2; 2) the structure shown in Fig.3 was placed in Zone 1 and all the corresponding RIRs were estimated; 3) the structure was moved to Zone 2 and all the corresponding RIRs were estimated. This process was repeated for all designated loudspeaker positions.

2.2. High-quality (HQ) RIR Measurements

In this setup, high-quality instrumentation has been used to estimate the RIRs. The microphone array is formed by Brüel&Kjaer type 4958 transducers, whereas JBL LSR305 loudspeakers, which present an accurate flat frequency-

response, were employed, as shown in Fig.2. For this setup, the 16 microphones (shown in Fig.4), and the 8 loudspeakers were connected to a professional soundcard. They were perfectly synchronized, which allowed to compute all the RIRs simultaneously. For this purpose, a specific Matlab application have been developed, effectively reducing time and error in the measurement process.

Figure 4. Rectangular array with Brüel&Kjaer type 4958 microphones.

2.3. Hybrid RIR Measurements

This setup is denoted as "hybrid configuration" because it uses the array of 16 Brüel&Kjaer microphones shown in Fig.4 connected by wire (through a soundcard) with the Yamaha NX-P100 loudspeaker. The 16 microphone positions for each zone were simultaneously estimated with the Matlab software as well.

3. DATA PREPROCESSING

The RIRs measured using the BT (wireless) connection suffered from a random delay inherent to the Bluetooth protocol. Although the Bluetooth standardization body has done a great effort to reduce its latency [4], the major drawback in Bluetooth communications is that each time a transmission starts, the latency, or delay, is random. To show how serious this problem is, consider a first transmission with an initial latency of 20 ms. Then stops transmitting and starts again a second transmission with an initial latency of 25 ms. This small difference of 5 ms is equivalent to moving the speaker 1,72 m apart from the microphone position, assuming the speed of sound being 345 m/s.

Therefore, we had to compensate the random delay introduced by the BT connection to preserve the coherence among the measured RIRs. For this purpose, we proposed two different pre-processing techniques.

3.1. Ideal delay computation based on the cross-correlation

Since a set of RIRs that have been estimated using a soundcard with almost perfect synchronization is available, we can compute any real delay between two different microphone positions given a fixed loudspeaker, or alternatively between two loudspeakers positions, given a fixed microphone.

Therefore, we firstly state one of the RIRs as the reference RIR, and secondly, we compute the relative delay (1) of the rest of the RIRs with respect to the reference by means of

$$\tau_{ij} = \max_{\tau} \frac{1}{2N+1} \sum_{\tau=-N}^{N} h_{ij}(n)\, h_{ref}(n+\tau)\,, \qquad (1)$$

where $h_{ref}(n)$ is the reference RIR, $h_{ij}(n)$ is the estimated RIR between the *i*th microphone and the *j*th loudspeaker and τ_{ij} is the relative delay calculated as the position (τ) of the maximum of their estimated cross-correlation function.

3.2. Ideal delay computation based on the devices' location

As shown in Fig.1, precise measurements of the location of the devices within the room are available. The reference RIR has been taken as the one between the first loudspeaker at the left side of the array and the microphone of Zone 1 located at the left upper corner of the array in Fig.1. The relative delay is calculated as an iterative process starting from the reference loudspeaker numbered as #1, calculating the difference between two distances: between speaker #1 to microphone #1 (reference) and between speaker #1 to microphone #*i*. The speed of the sound divided by this difference gives the relative delay τ_{i1}. This process is repeated with the rest of loudspeakers and all the microphones till all delays are computed.

3.3. Delay compensation

As a final step, the real random delays between the RIRs measured by the low-cost system of Section 2.1 are calculated using (1) and then compensated by adding or dropping as many zero samples as required to the measured RIRs, such that the resulting RIRs match the ideal delay computed above.

3.4. Comparison of estimated RIRs

Fig.5 shows an example of the estimated RIR between the 4th loudspeaker and the 6th microphone for each zone. Left side corresponds to "Zone 1" and right side to "Zone 2". It can be noticed that the RIRs obtained by the low-cost devices present a different delay, whereas the other two sets of RIRs are synchronized. To further investigate the differences between the three collections, Fig.6 shows the absolute value

of the $h_{ij}(n)$ of Zone 1 shown in the left side of Fig.5. The *y* axis is depicted in logarithmic scale.

The RIR estimated with the tablet presents an abrupt fall compared to the other two RIRs, and its energy remains on a constant level of -20 dB along the time. This behavior indicates that the microphone of the tablet introduces significant noise in the decorrelation of the recorded signal. Regarding the differences between the RIR obtained by the hybrid system and the high-quality (HQ) system, Fig.5-6 show a lower density of captured reflections in HQ.

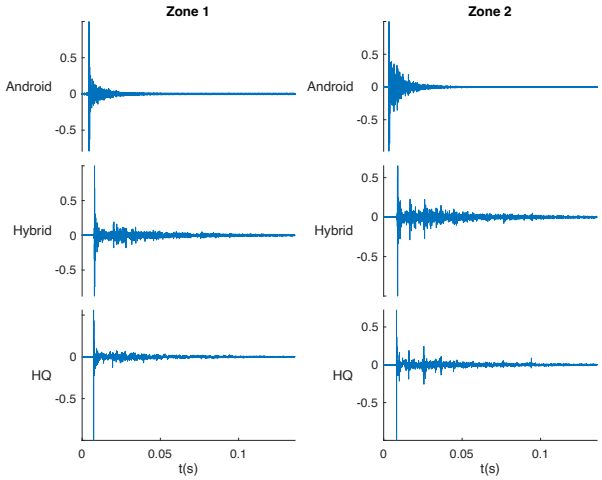


Figure 5. Measured RIRs between the 4th loudspeaker and the 6th microphone of each zone. Left side corresponds to "Zone 1" and right side to "Zone 2".

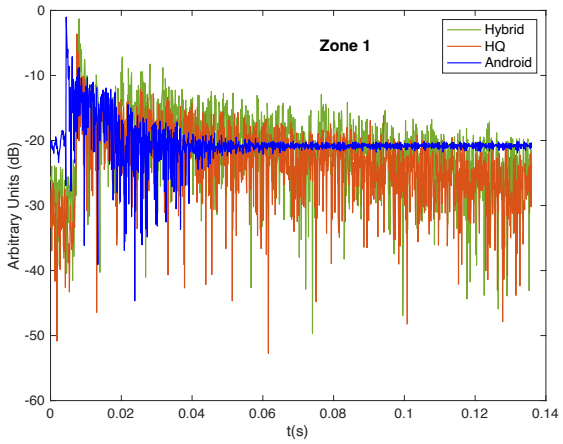


Figure 6. Logarithmic magnitude of the RIRs of Zone 1 shown in the left side of Fig.5.

4. EXPERIMENTAL RESULTS

In this section, we will compare the performance of a PSZ system when the related RIRs have been estimated via one of the combinations of microphones and loudspeakers described in Section 2. Since for the case of the Android device, the delays have been compensated using two different approaches, a total of four different sets of RIRs will be compared.

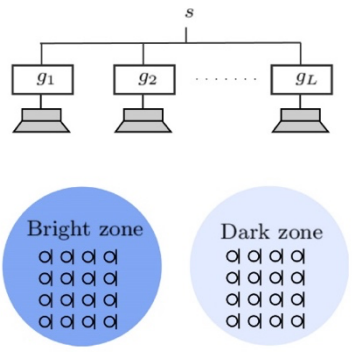


Figure 7. Personal Sound Zones system.

4.1. Personal Sound Zones

As shown in Fig.7, a PSZ system is designed such that it delivers the sound *s* to a zone, called the "bright zone", and tries to cancel it in another zone, called the "dark zone". For this purpose, a set of *L* finite impulse response (FIR) filters, one per loudspeaker, is designed according to the following optimization function [5]:

$$\mathbf{g} = \min_{\mathbf{g}} \{ \underbrace{\mathbf{x}_d^T \mathbf{x}_d}_{\text{Energy dark}} + \underbrace{||\mathbf{x}_b - \mathbf{d}_b||^2}_{\text{Error bright}} + \lambda \underbrace{\mathbf{g}^T \mathbf{g}}_{\text{Filter's Energy}} \} \qquad (2)$$

where $\mathbf{g}$ is the matrix of the filters of dimensions L_f x L, being L_f the length of the FIR filters and L the number of loudspeakers, λ is a regularization parameter, $\mathbf{d}_b$ is the desired response in the bright zone (usually the RIR from one of the central loudspeakers), and $\mathbf{x}_b$ and $\mathbf{x}_d$ are the combined acoustic responses at the bright and dark zones respectively, that is, $\mathbf{x}_b = \mathbf{g} * \mathbf{h}_b$ and $\mathbf{x}_d = \mathbf{g} * \mathbf{h}_d$, where $\mathbf{h}_b$ and $\mathbf{h}_d$ are the set of RIRs between the loudspeaker array and the microphone array of the bright and dark zones, respectively.

4.2. Filter computation in the PSZ system

As stated above, the design of the PSZ filters depends on the set of estimated RIRs ($\mathbf{h}_b$, $\mathbf{h}_d$) implicitly used in (2). Moreover, the set of filters comprises two subsets: one set of filters must be designed for the Zone 1 acting as the bright zone and Zone 2 as the dark one, which we denote by $\mathbf{g}_1$, and another set of filters, $\mathbf{g}_2$, is designed for the reverse condition where Zone 2 is the bright zone and Zone 1 the dark one.

Consequently, two sets of filters have been computed for each collection of RIRs described in Sections 2 and 3.

4.3. Objective comparison: acoustic contrast

The acoustic contrast (AC) is the most used metric to assess the performance of a PSZ system. It is defined as [6]:

$$AC = \frac{\mathrm{E}_b}{\mathrm{E}_d} = \frac{E[x_b^T x_b]}{E[x_d^T x_d]} \qquad (3)$$

where E_b and E_d are the average energy gains that would enhance (or attenuate) the sound s of Fig.7 at the bright and dark zones, respectively. The AC can be computed in frequency for every bin or, alternatively, by averaging 1/3-octave bands to improve the readability of the results [7].

To perform a fair comparison among the different collections of RIRs and considering that a personal sound zone is intended to give a perceptual experience of the sound at that location, we have measured an additional set of binaural RIRs at each zone using a Neumann KU100 dummy head. Therefore, the AC (3) is computed using the set of filters for each RIR collection, but using the binaural RIRs to compute the combined acoustic responses **x**:

$$\mathrm{x}_z^{\mathrm{set}} = \mathrm{g}^{\mathrm{set}} * \mathrm{h}_{\mathrm{binaural},z} \qquad (4)$$

where "set" denotes the RIR collection and z can be b (bright) or d (dark).

Fig.8 and Fig.9 show the averaged AC in frequency when Zone 1 and Zone 2 are the bright zone, respectively. The frequency axis is depicted in logarithmic scale in the range [125, 2000] Hz, since the loudspeaker separation introduces aliasing above 1kHz. It can be noticed that Zone 1 (Fig.8) presents higher AC values at very low frequencies than Zone 2 (Fig.9), but from 200 Hz on, both zones achieve similar contrast. Regarding the comparison between RIRs measurements, the filters computed from the RIRs obtained with the high-quality instrumentation ("HQ") show, in average, AC values 5 dB above those from the rest of sets. The filters computed from the RIRs obtained through the "Hybrid" set present a good behavior in very low frequencies but achieve a poor AC value of 5 dB in the range of 1kHz, where our hearing system is more sensitive.

Finally, the filters computed from the RIRs obtained by the tablet and the BT loudspeaker ("Android") behave very similar independently if the random delays were compensated by computing the cross-correlation ("corr") or from their locations ("theo") (see Section 3). Both present a poor behavior at very low frequencies, but obtain good contrast around the most sensitive frequencies, outperforming the "Hybrid" set. Therefore, we can conclude that the RIRs obtained by the low-cost devices can achieve good AC values when used in the design of PSZ systems.

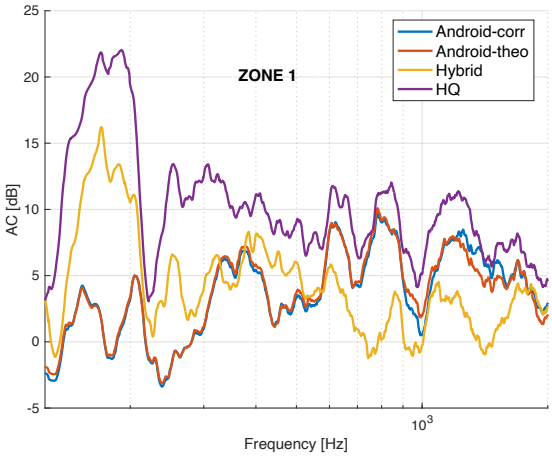


Figure 8. Acoustic Contrast in Zone 1.

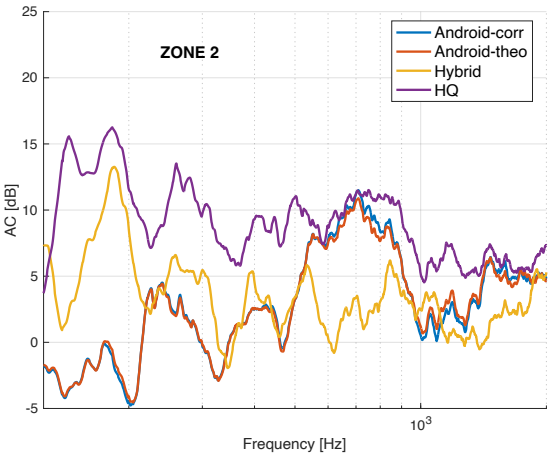


Figure 9. Acoustic Contrast in Zone 2.

4.4. Subjective comparison: Psychoacoustic test

A subjective test has been carried out to assess the performance of the four RIR collections by evaluating the audio quality delivered to the zones by the PSZ system. We have added a fifth profile where no filter is used, that is, simulating that the PSZ system is off.

The stimuli of the test consisted in two speech signals, one male voice and one female voice, both recorded in an anechoic chamber and speaking Spanish. The duration of each stimulus was 8 s and the sampling frequency was f_s=44100 Hz. They were generated such that Zone 1 would be the bright zone for the male voice and Zone 2 the bright zone for the female one, using the combined acoustic response in (4) for each set and zone. The test was carried out by 14 participants for Zone 1 and 15 participants for Zone 2. All of them indicated no hearing loss and their repetibility and

consistency were considered valid (thresholds of 50% for repetibility and 75% for consistency).

The test has been carried out using the Two-Alternative Forced Choice (2-AFC) protocol, where two different stimuli (audio signals) were presented to the assessors, and they had to choose one of them according to the following question: "*Which of the two audio signals has the least interfering voice?*". Therefore, the participants evaluated which of the audios were perceived with minimum interference, thus, with better quality.

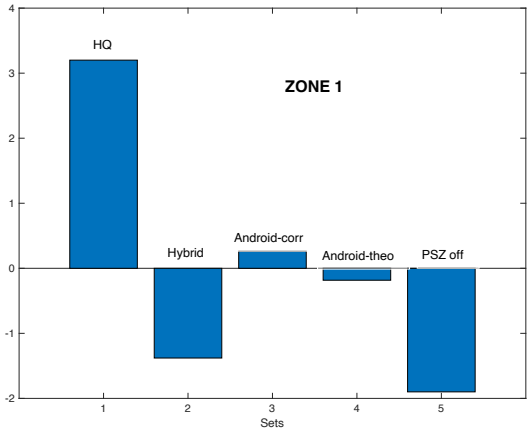


Figure 10. Preference for the audios obtained in Zone 1.

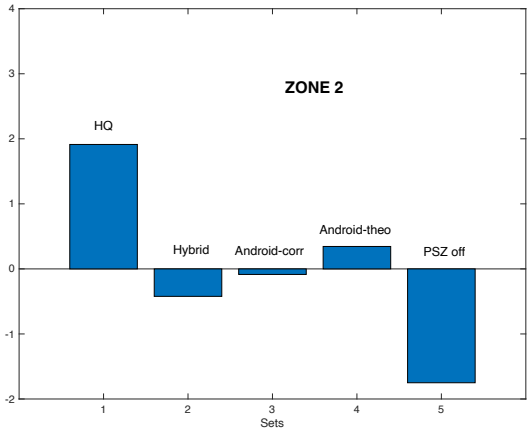


Figure 11. Preference for the audios obtained in Zone 2.

Fig.10 and Fig.11 show the value of merit (VoM) of each of the audio signals presented in the jury test. The VoM of an audio signal relates to the number of times that the audio has been chosen compared to the rest of signals. Another significant characteristic is that the sum of all the values of merit is 1. Therefore, if an audio signal has a positive VoM means that it has been chosen more often that other with a negative VoM, but if the range of the minimum VoM to the maximum VoM is wide, it means that most of the participants agree with that selection. In this sense, the range of VoM is larger for Zone 1 (Fig.10) compared to that exhibited by Zone 2 (Fig.11), that is, most juries would agree on the results shown for Zone 1. Regarding the comparison among the four PSZ designs, "HQ" obtained the best perceived audio quality. Surprisingly, the PSZ system that uses the "Hybrid" RIRs is almost comparable to doing nothing in Zone 1, although in Zone 2 performs like both "Android" collections. This behavior can be related to the AC shown in Fig.8-9 by the "Hybrid" set, which was much lower than the ACs of the "Android" and "HQ" sets around the sensitive band of 1kHz.

Therefore, we can state that estimating RIRs with smartphones or tablets connected to Bluetooth loudspeakers, can be a rough but cost-efficient alternative to the use of professional instrumentation when designing PSZ systems.

5. CONCLUSIONS

The estimation of the room impulse responses (RIRs) required to design a personal sound zones (PSZ) system are usually obtained using professional/high-quality equipment, which is expensive and difficult to relocate. In this work, we have proposed an alternative low-cost system comprising an Android tablet and a Bluetooth (BT) loudspeaker to estimate the desired RIRs. We have studied the performance of both systems, together with a hybrid approach, when used to design a PSZ system. The evaluation has been carried out using an objective metric (acoustic contrast) and a subjective psychoacoustic test. Surprisingly, the low-cost system performs better than the hybrid one and can be a realistic cost-efficient alternative to professional instrumentation.

6. REFERENCES

[1] T. Betlehem, W. Zhang, M. A. Poletti and T. D. Abhayapala, "Personal Sound Zones: Delivering interface-free audio to multiple listeners," in *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 81-91, March 2015.

[2] S. J. Elliott, J. Cheer, J. -W. Choi and Y. Kim, "Robustness and Regularization of Personal Audio Systems," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 7, pp. 2123-2133, Sept. 2012.

[3] G. B. Stan, J. J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," J. Audio Eng. Soc., vol. 50, no. 4, pp. 249-262, 2002.

[4] Aza, A., Melendi, D., García, R. *et al.* "Bluetooth 5 performance analysis for inter-vehicular communications.", *Wireless Netw*, 28, 137–159, 2022.

[5] Moles-Cases, V., Piñero, G., de Diego, M., Gonzalez, A., "Personal Sound Zones by Subband Filtering and Time Domain Optimization". *IEEE/ACM Tran. on Audio, Speech, and Lang. Proc.*, *28*, 2684–2696, 2020.

[6] Choi, J.-W., Kim, Y.-H., "Generation of an acoustically bright zone with an illuminated region using multiple sources." *J. Acoust. Soc. Am.* 111 (4): 1695–1700, 2002.

[7] P. D. Hatziantoniou, J. N. Mourjopoulos, "Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses,", *J. Acoust. Soc. Am.* 48 (4): 259-280, 2000.