TECNIACUSTICA ELCHE · 2022

53º CONGRESO ESPAÑOL DE ACÚSTICA
XII CONGRESO IBÉRICO DE ACÚSTICA

Implementation of a Dynamic Personal Sound Zones System

PACS: 43.60.Ac, 43.60.Gk, 43.60.Hj

Molés-Cases, Vicent

Universitat Politècnica de València, Camí de Vera, s/n, 46022 València, España, 96389580, vimoca3@iteam.upv.es

Fuster, Laura

Universitat Politècnica de València, Camí de Vera, s/n, 46022 València, España, 96389580, lfuster@iteam.upv.es

Piñero, Gema

Universitat Politècnica de València, Camí de Vera, s/n, 46022 València, España, 96389580, gpinyero@iteam.upv.es

Gonzalez, Alberto

Universitat Politècnica de València, Camí de Vera, s/n, 46022 València, España, 96389580, agonzal@dcom.upv.es

Palabras Clave: personal sound zones, tracking, multichannel processing

ABSTRACT

Personal Sound Zones (PSZ) systems aim to render different audio signals to multiple listeners within a room, such that two persons located in different places of the room can listen different audio programs without the need of headphones. PSZ systems commonly use an array of loudspeakers combined with signal processing techniques. In this work, we describe an implementation of a dynamic PSZ system, in which the listeners can move within the room and the system continues to provide them their respective audio signals. In this way, the PSZ system must update the locations of the listeners, together with the signals fed to the array, to properly enhance the audio program of interest to each listener while cancelling the interference between them. The proposed implementation uses multichannel signal processing techniques together with a 3D camera to determine the actual position of the listeners. This PSZ system achieves a level of isolation between reproduction zones of 10dB in the frequency range 100-6000Hz.

1 INTRODUCTION

Personal Sound Zones (PSZ) systems aim to deliver different sounds to a number of users in a shared space by using arrays of loudspeakers [1]. To achieve this, a set of filters is used to process the audio signals that are fed to the loudspeakers. Different techniques have been proposed to compute the filters, and among these, Acoustic Contrast Control (ACC) [2] is the algorithm that can achieve highest isolation between the bright and dark zones, where the terms bright and dark zone refer to the regions where we want high and low acoustic energy, respectively [2]. However, ACC cannot synthesize a specific target response in the bright zone. To solve this limitation, the weighted Pressure Matching (wPM) algorithm was proposed [3]. It offers the possibility to render a target response in the bright zone while keeping control over the energy in the dark zone. Three different formulations of wPM, each offering some advantages and disadvantages, can be used to compute the filters of the PSZ system, namely, the frequency domain formulation (wPM-F) [3], the time-domain formulation (wPM-T) [4], and the subband-domain formulation (wPM-S) [5]. In particular, the wPM-S algorithm is a good option for dynamic PSZ systems, since it offers a good balance of performance, computational complexity, and latency. Furthermore, the wPM-S is a good choice for PSZ systems using several arrays of loudspeakers, since it allows each array to operate only in the subbands in which it provides good directivity (without requiring the use of cross-over filters).

In this work, we describe a practical implementation of a dynamic PSZ system in which two users can move within the room and the system keeps providing them their respective audio signal. The system uses two arrays of loudspeakers to generate the sound zones, and a 3D camera to track the position of the users. The filters of the system are updated every 200ms and are calculated by wPM-S algorithm.

The outline of the paper is as follows. Section 2 reviews the wPM-S algorithm and Section 3 presents the proposed implementation. Section 4 evaluates the level of acoustic isolation that can be achieved between listeners. Finally, Section 5 summarizes the main conclusions.

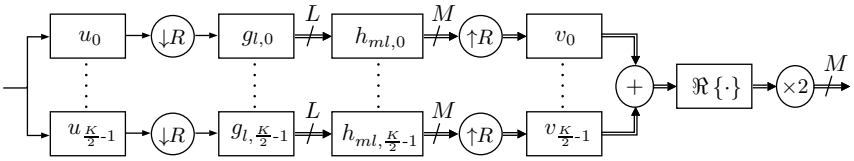


Figure 1. Model for computing the subband filters of a PSZ system using the wPM-S algorithm.

2 WEIGHTED PRESSURE MATCHING WITH SUBBAND DOMAIN FORMULATION

Let us define L as the number of loudspeakers of the PSZ system, M as the number of control points, and h_{ml} as the Room Impulse Response (RIR) between the l-th loudspeaker and the m-th control point. The system model shown in Figure 1, which employs a Generalized Discrete Fourier (GDFT) filter bank with K subbands and a resampling factor R, is proposed in [5] for PSZ systems. For GDFT filter banks, the analysis filters $u_k(n)$ are obtained by modulating an I_p-length low-pass prototype filter $p(n)$ and the synthesis filters $v_k(n)$ are time-reversed and conjugated versions of the analysis filters, i.e.,

$$u_k(n) = p(n)e^{j\frac{2\pi}{K}\left(k+\frac{1}{2}\right)n}, \tag{1}$$

$$v_k(n) = u_k^*(I_\mathrm{p} - 1 - n). \tag{2}$$

The model in Figure 1 includes in each subband the FIR subband filters $g_{l,k}$ of length I_g, and the subband components of the RIR $h_{ml,k}$ of length I_h. The subband components $h_{ml,k}$ are obtained by applying the subband decomposition algorithm proposed in [6] to the RIRs h_{ml}. Now, let us define the cascade impulse response in the m-th control point and the k-th subband as

$$x_{m,k}(n) = \sum_{l=0}^{L-1} h_{ml,k}(n) * g_{l,k}(n). \tag{3}$$

Also, we can define a vector of $x_{m,k}(n)$ in all time instants, all control points, and subband k as

$$\mathbf{x}_k = \mathbf{H}_k\mathbf{g}_k, \tag{4}$$

where $\mathbf{g}_k = [\mathbf{g}_{0,k} \quad \cdots \quad \mathbf{g}_{I_\mathrm{g}-1,k}]^T$ and $\mathbf{g}_{n,k} = [g_{0,k}(n) \quad \cdots \quad g_{L-1,k}(n)]^T$, and in which

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{0,k}^T & \cdots & \mathbf{H}_{I_h-1,k}^T & \mathbf{0}_{LxM} & \cdots & \mathbf{0}_{LxM} \\ \mathbf{0}_{LxM} & & & & & \\ \vdots & & Toeplitz & & & \\ \mathbf{0}_{LxM} & & & & & \end{bmatrix}^T, \tag{5}$$

where

$$\mathbf{H}_{n,k} = \begin{bmatrix} h_{00,k}(\mathrm{n}) & \cdots & h_{0(L-1),k}(\mathrm{n}) \\ \vdots & \ddots & \vdots \\ h_{(M-1)0,k}(\mathrm{n}) & \cdots & h_{(M-1)(L-1),k}(\mathrm{n}) \end{bmatrix}. \tag{6}$$

Now, let us define the target response $d_{m,k}$ of length $I_\mathrm{d} = I_\mathrm{h} + I_\mathrm{g} - 1$ for the m-th control point as [7]

$$d_{m,k}(n) = \begin{cases} h_{ml_\mathrm{r},k}(n - \tau_\mathrm{d}) & m \in \mathcal{B} \\ 0 & m \in D \end{cases} \tag{7}$$

where $\mathcal{B}$ and D are the sets of the indices of the control points in the bright and dark zones, respectively, l_r is the index of the reference loudspeaker, and τ_d is a modelling delay that assures the causality of the filters. Now, let us define a column vector containing the target response for all time instants, all control

points, and subband k as $\mathbf{d}_k = [\mathbf{d}_{0,k} \quad \dots \quad \mathbf{d}_{I_d,k}]^T$, where $\mathbf{d}_{n,k} = [d_{0,k}(n) \quad \dots \quad d_{M-1,k}(n)]^T$. Then, the optimal subband filters in the k-th for the wPM-S algorithm are given by

$$\mathbf{g}_{\mathrm{opt},k} = \underset{\mathbf{g}_k}{\operatorname{argmin}}\{\|\mathbf{W}_k(\mathbf{H}_k\mathbf{g}_k - \mathbf{d}_k)\|^2 + \beta\|\mathbf{g}_k\|^2\} = (\mathbf{H}_k^H\mathbf{W}_k^T\mathbf{W}_k\mathbf{H}_k + \beta\mathbf{I})^{-1}\mathbf{H}_k^H\mathbf{W}_k^T\mathbf{W}_k\mathbf{d}_k, \tag{8}$$

where $\beta > 0$ is a regularization factor, and $\mathbf{W}_k$ is a diagonal weighting matrix for the k-th subband, whose i-th diagonal element is defined as

$$[\mathbf{W}_k]_{(i,i)} = \begin{cases} \sqrt{u_k} & (i \bmod M) \in \mathcal{B} \\ \sqrt{1-u_k} & (i \bmod M) \in D \end{cases}, \tag{9}$$

in which $0 \leq u_k \leq 1$ is a weighting factor.

3 DESCRIPTION OF THE IMPLEMENTED SYSTEM

The implemented system was used to provide individualized sound zones to two users that change their position over time in an office-like room of size 7.2 x 11.72 x 2.65m. The room presents a reverberation time of $T_{60} = 500$ms. It is considered that both users can be located in any of the 5 zones shown in Figure 2. The system is formed by two subsystems, as shown in Figure 3. On the one hand, the tracking subsystem, which determines the position of the users in each time instant. On the other hand, the audio processing subsystem, which processes the signals that are fed to the loudspeakers to generate the sound zones in the required locations. Both subsystems include a computer, and UDP datagrams are used for the communication between both subsystems.

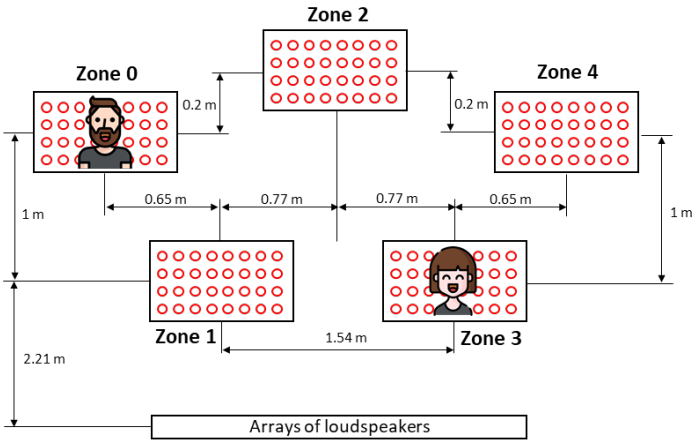


Figure 2. Zones in which the two users of the system can be located.

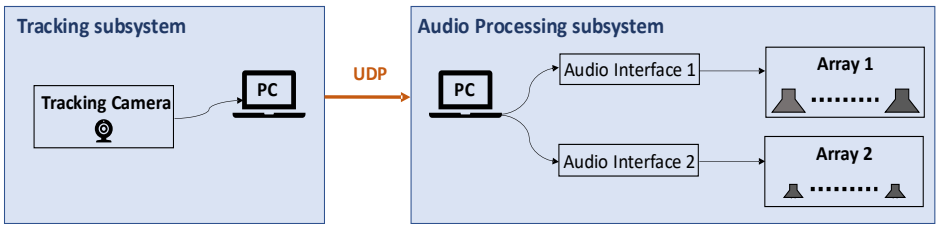


Figure 3. Diagram of the proposed dynamic PSZ system.

3.1 Tracking subsystem

The main function of this subsystem is to track the position of the two users of the system. Specifically, this subsystem is formed by an Orbbec Astra Series 3D camera [8] which is connected to a computer MSI Modern 14 A10RAS with Intel Core i7. The computer processes the images captured by the camera and determines the position of the two users every 200 ms. The images are processed using code implemented using the Orbbec Body Tracking SDK [9] for C language. The estimated position for user

i is a vector defined as $\mathbf{u}_i = \begin{bmatrix} u_i^x & u_i^y & u_i^z \end{bmatrix}$. The position $\mathbf{u}_i$ of the two users is estimated every 200 ms and is sent to the Audio Processing subsystem using UDP datagrams.

Figure 4. Picture showing the array of woofers, the array of tweeters and the 3D camera.

3.2 Audio Processing subsystem

The main function of this subsystem is to generate the sound zones in the current locations of the users.

3.2.1 Hardware

This subsystem uses a computer HP EliteDesk 800 with Intel Core i7, in which MATLAB 2018b is used to process the signals that are transferred to two audio interfaces Roland Studio-Capture 1610 [10], whose outputs are then fed to two arrays of loudspeakers. In particular, an array of 8 woofers JBL 305P MkII [11] and an array of 8 tweeters Visaton K23 [12] are used, as shown in Figure 4. We use two arrays of loudspeakers to obtain good directivity in a broadband frequency range. In particular, the array of woofers, which presents an inter-element distance of 18 cm, can provide good directivity in the frequency range 100-1500Hz, but for higher frequencies the effect of the spatial aliasing degrades the performance. On the other hand, the array of tweeters, with an inter-element distance of 5 cm, can provide good directivity within the frequency range 1000-6000Hz. Thus, the combination of the two arrays can provide good directivity in the broadband frequency range 100-6000Hz. We used the exponential swept-sine technique [13] to measure the RIRs $h_{ml,k}$ between all the loudspeakers of the system and a grid of 4x8 Bruël & Kjær microphones Type 4958 [14] (shown in Figure 5), which was located in each of the 5 considered zones. The grid of microphones was previously calibrated using a Bruël & Kjær sound calibrator Type 4231 [15].

Figure 5. Grid of 4x8 Bruël & Kjær microphones Type 4958.

3.2.2 Signal processing

The signal processing blocks applied by the Audio Processing subsystem, which have been implemented using MATLAB 2018b and operate at a sampling frequency of 14700Hz, are shown in Figure 6. Let us denote s_i as the audio signal that we want to deliver to the i-th user of the system. For each user i, the input signal s_i is processed by the analysis section of a GDFT filter bank with $K = 30$ subbands, resampling factor $R = 22$, and prototype filter of length $I_\mathrm{p} = 145$. We show the spectrum of the analysis filters in the positive subbands of the filter bank in Figure 7. Once the subband signals at the output of the analysis section are obtained, the subband signals for the i-th user are filtered using the filters $g_{l,k}^{(b_i,d_i)}$. The filters $g_{l,k}^{(b_i,d_i)}$ are computed using wPM-S, where b_i and d_i indicate the indices of

the zones considered as the bright and dark zones, respectively, for computing the filters for user i. Once the signals for each user are filtered using the subband filters, these signals are combined before being fed to the synthesis stage of the GDFT filter bank. Finally, the signals at the output of the filter bank are transferred to the audio interfaces, and later, the signals at the output of the audio interfaces are fed to the arrays of loudspeakers. Finally, it is important to note that the analysis and synthesis filtering of the GDFT filter bank is performed using the computationally efficient polyphase implementation proposed in [16], and that the filtering in the subbands is performed using the overlap-and-save method [17].

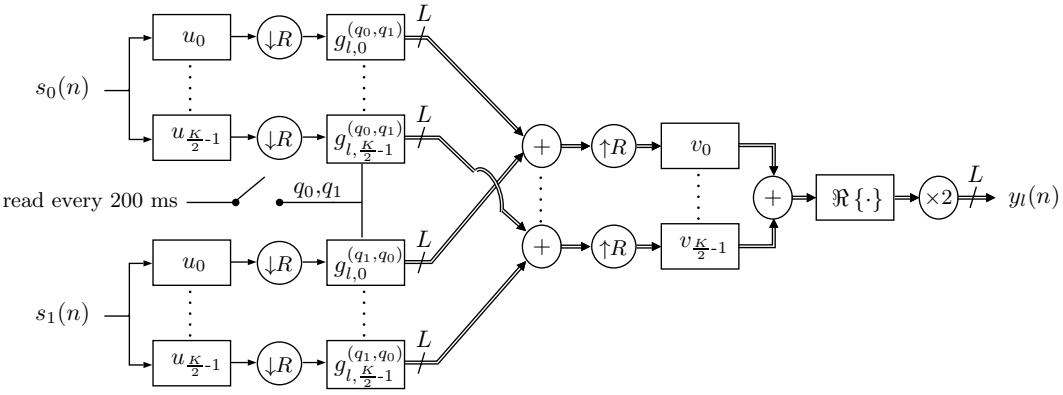


Figure 6.Signal processing blocks applied by the Audio Processing subsystem.

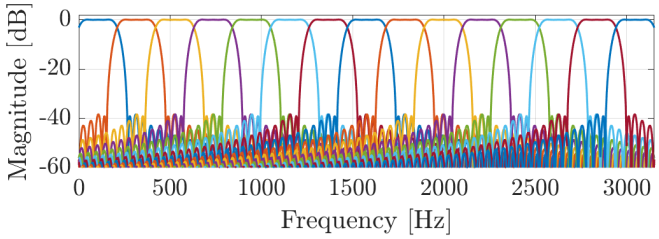


Figure 7. Spectrum of the analysis filters in the positive subbands for the considered filter bank.

3.2.3 Filter computation

An important aspect is that the filters $g_{l,k}^{(b_i,d_i)}$ are not computed in real-time. Instead, the wPM-S filters for all the 20 possible combinations of bright and dark zones in the system are computed offline using (8). The filters are computed using the functionalities provided in [18], with a filter length $I_{\mathrm{g}} = 180$, a modelling delay $\tau_{\mathrm{d}} = 90$, and a regularization factor $\beta = 10^{-3}$. The weighting factor in (9) is selected as $u_k = 0.5$ for subbands $k = 0$ to $k = 3$ and as $u_k = 0.9$ for subbands $k = 4$ to $k = 14$, because we observed that more effort is needed in the optimization to minimize the errors in the dark zone than in the bright zone for mid and high frequencies. Moreover, each array of loudspeakers only operates in the subbands in which can provide good directivity, i.e., in subbands $k = 0$ to $k = 3$ and subbands $k = 4$ to $k = 14$ for the arrays of woofers and tweeters, respectively. Thus, we set to 0 the filters $g_{l,k}^{(b_i,d_i)}$ in subbands $k = 4$ to $k = 14$ and loudspeakers $l = 8$ to $l = 15$, i.e., the array of tweeters, and in subbands $k = 0$ to $k = 3$ and loudspeakers $l = 0$ to $l = 7$, i.e., the array of woofers.

3.2.4 Filter update

The subband filters $g_{l,k}^{(b_i,d_i)}$ for each user are updated every 200ms by following the procedure described next. First, the position $\mathbf{u}_i$ for the two users is read from the UDP datagrams. Then, the index q_i of the zone to which user i is closer is determined as:

$$q_i = \underset{q}{\operatorname{argmin}} \left\{ \left\| \mathbf{u}_i - \mathbf{z}_q \right\|^2 \right\}, \tag{10}$$

where $\mathbf{z}_q = [z_q^x \quad r_q^y \quad r_q^z]$ are the coordinates of the centre of zone q. Specifically, for user 0 we consider that the bright zone is $b_0 = q_0$ and the dark zone $d_0 = q_1$, while for user 1 we consider $b_1 = q_1$ and $d_1 = q_0$. Once the indices of the zones q_i are known, the filters are updated by finding the set of pre-computed filters that generate the sound zones in the locations determined by the indices q_i.

4 EVALUATION RESULTS

Next, we present experimental results to evaluate the performance of the implemented system. In particular, we evaluate the performance of the system using: 1) the Acoustic Contrast (AC), which is the ratio between the mean energies produced by the system in the control points of the bright and dark zones; and 2) the Mean Squared Error (MSE) in the control points of the bright zone with respect to the selected target response. The AC is related to the level of acoustic isolation between users and the MSE is related to the reproduction errors in the bright zone.

We show in Figure 8 and Figure 9 the AC and the MSE, respectively, for different combinations of bright and dark zones for the considered system. In general, we can see that the AC is above 10dB in the frequency range 100-6000Hz for all the considered combinations. This means that each user listens the sound provided to the other user attenuated at least by 10dB. Also, we can see that the MSE is below 0dB in the frequency range 100-6000Hz for all the considered combinations. In this case, an important aspect is that frequencies above 1500 Hz present higher MSE than the lower frequencies. This is because in these frequencies we selected a weighting factor of $u_k = 0.9$, which leads to AC levels above 10dB at the cost of worsening the MSE in the bright zone. The motivation for trading off the level of MSE by AC in the frequencies above 1500Hz is that in these frequencies the human perception of sound is more sensitive to intensity differences than to phase differences [19]. Then, obtaining lower interference level for the high frequencies is a better perceptual option than obtaining an accurate synthetization of the target response in the bright zone. Moreover, it is interesting to note that there are important differences between the AC levels that can be achieved for the different combinations of bright and dark zones. In particular, for $b_i = 0$ and $d_i = 4$, and for $b_i = 2$ and $d_i = 4$ we can achieve an AC higher than 15dB in the frequency range 100-1000Hz, while for other combinations such high levels of AC cannot be obtained. This is may be caused by the distance between the bright and dark zones, since usually the longer the distance is between the zones the larger the isolation that can be achieved between them. Finally, we illustrate the operation of the dynamic PSZ system in a video that can be found in [20].

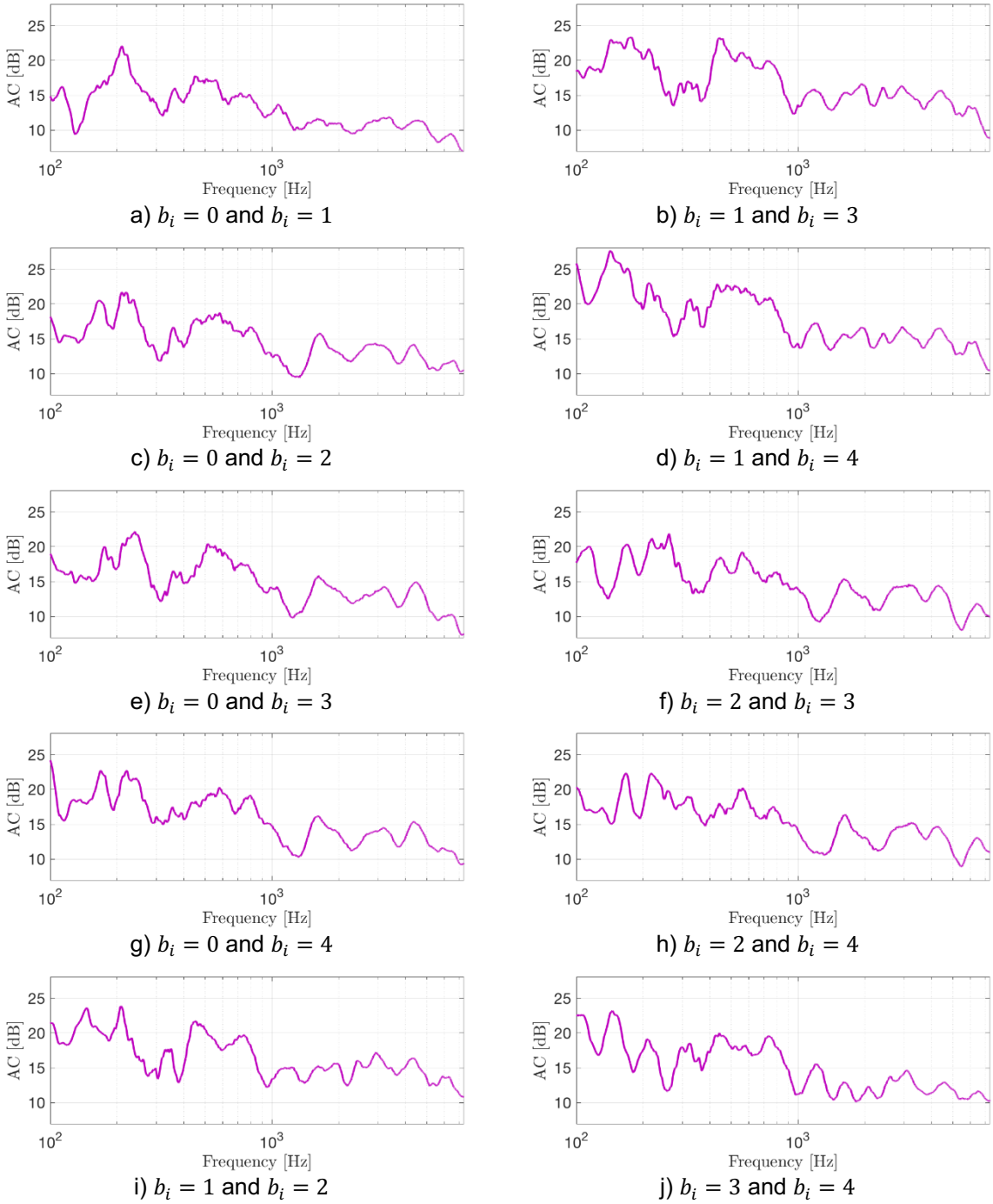
5 CONCLUSIONS

In this work we described a practical implementation of a dynamic PSZ, in which the position of two users can change over time. The implemented system consists of two subsystems. On the one hand, the tracking subsystem, which uses a 3D camera connected to a computer to track the position of the users. On the other hand, the audio processing subsystem, which uses a computer, an array of 8 woofers and an array of 8 tweeters to generate sound zones in the positions where the users of the system are located. The tracking subsystem sends the position of the users every 200ms to the audio processing subsystem using UDP datagrams. The audio processing subsystem, based on the position of the users, selects from a set of pre-computed filters those that can generate sound zones in the current location of the users. The filters of the system are computed using the wPM-S algorithm. An important aspect is that each of the arrays of loudspeakers only operates in the frequency range in which can provide good directivity. In particular, the array of woofers operates in the range 100-1500Hz, and the array of tweeters in the range 1000-6000Hz. We presented experimental evaluation results that show that the implemented system can achieve an Acoustic Contrast higher than 10dB in the frequency range 100-6000Hz for all the studied combinations of locations for the bright and dark zones.

ACKNOWLEDGEMENTS

The authors would like to thank Javier García Morant and Eric Beaucamps Santofimia for their valuable collaboration in the preparation and set up of the implemented PSZ system. Vicent Molés-Cases was supported by the Spanish Ministry of Education through Grant No. FPU17/01288. This research was

supported by the Spanish Ministry of Science, Innovation and Universities through Grant No. RTI2018-098085-B-C41 (MCIU/AEI/FEDER, UE) and PID2021-124280OB-C21 (MCIU/AEI/FEDER, UE).



a) $b_i = 0$ and $b_i = 1$ b) $b_i = 1$ and $b_i = 3$

c) $b_i = 0$ and $b_i = 2$ d) $b_i = 1$ and $b_i = 4$

e) $b_i = 0$ and $b_i = 3$ f) $b_i = 2$ and $b_i = 3$

g) $b_i = 0$ and $b_i = 4$ h) $b_i = 2$ and $b_i = 4$

i) $b_i = 1$ and $b_i = 2$ j) $b_i = 3$ and $b_i = 4$

Figure 8. Acoustic Contrast as a function of frequency for different combinations of bright zone and dark zone indices, i.e., b_i and d_i, respectively.

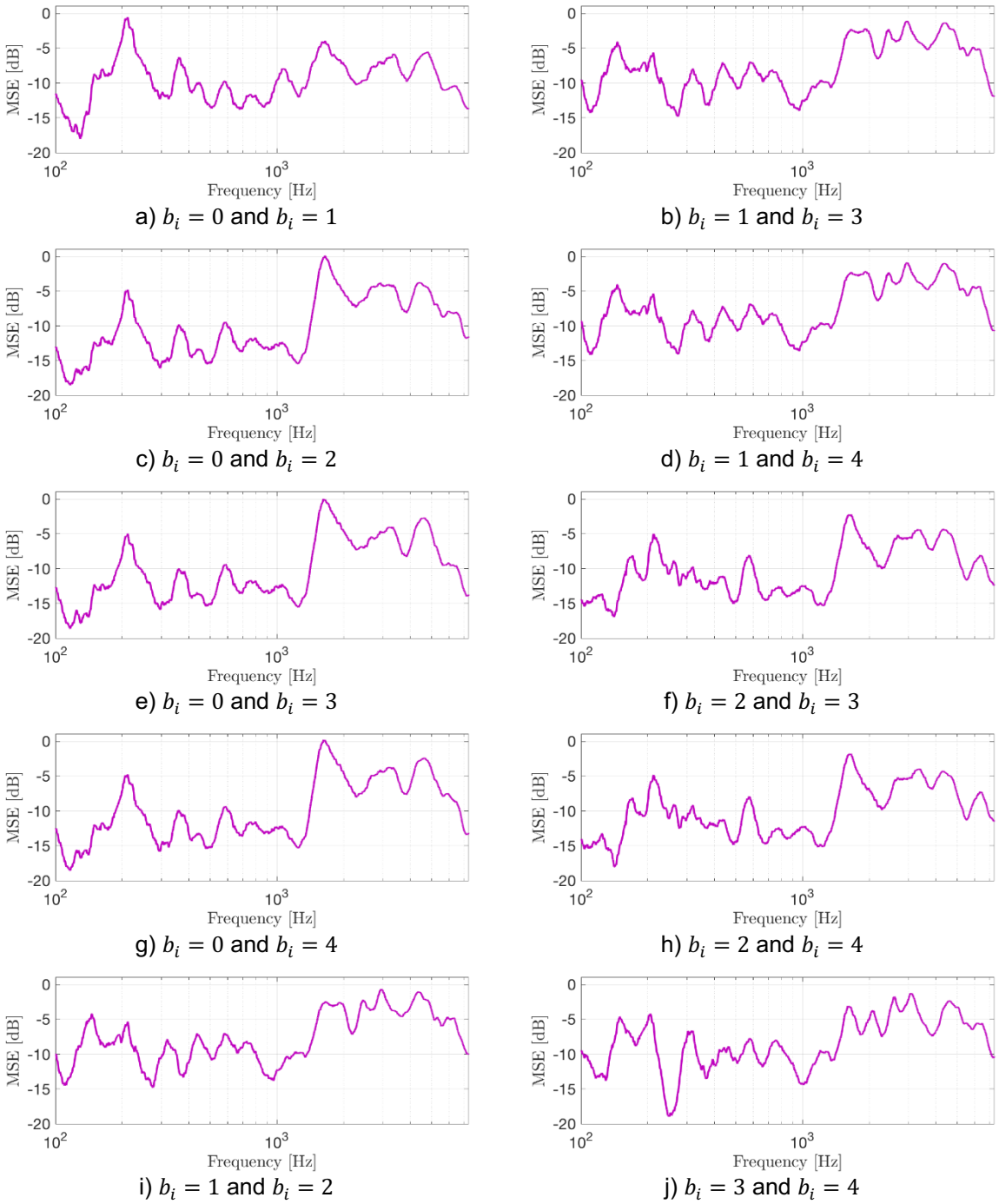


Figure 9. Mean Squared Error in the bright zone as a function of frequency for different combinations of bright zone and dark zone indices, i.e., b_i and d_i, respectively.

REFERENCES

[1] W. F. Druyvesteyn and R. M. Aarts, "Personal sound," *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 685–701, 1997, doi: 10.1121/1.410932.

[2] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region

using multiple sources," *J. Acoust. Soc. Am.*, vol. 111, p. 1695, 2002, doi: 10.1121/1.1456926.

[3] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, pp. 4518–4525, 2012, doi: 10.1121/1.4714349.

[4] M. F. Simon Galvez, S. J. Elliott, and J. Cheer, "Time Domain Optimization of Filters Used in a Loudspeaker Array for Personal Audio," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 1869–1878, Nov. 2015, doi: 10.1109/TASLP.2015.2456428.

[5] V. Molés-Cases, G. Pinero, M. De Diego, and A. Gonzalez, "Personal Sound Zones by Subband Filtering and Time Domain Optimization," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 2684–2696, 2020, doi: 10.1109/TASLP.2020.3023628.

[6] J. P. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, "The complex subband decomposition and its application to the decimation of large adaptive filtering problems," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2730–2743, 2002, doi: 10.1109/TSP.2002.804068.

[7] V. Molés-Cases, S. J. Elliott, J. Cheer, G. Piñero, and A. Gonzalez, "Weighted pressure matching with windowed targets for personal sound zones," *J. Acoust. Soc. Am.*, vol. 151, no. 1, p. 334, Jan. 2022, doi: 10.1121/10.0009275.

[8] "Astra Series-Orbbec 3D." http://orbbec3d.com/index/Product/info.html?cate=38&id=36

[9] "Orbbec Tracking SDK." http://orbbec3d.com/index/Product/info.html?cate=38&id=38

[10] "Roland STUDIO-CAPTURE 1610." https://www.roland.com/es-es/products/studio-capture/

[11] "JBL 305P MkII." https://jblpro.com/en-US/products/305p-mkii

[12] "Visaton K23." https://www.visaton.de/en/products/miniature-speakers/k-23-8-ohm

[13] Angelo Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES 108th conv, Paris, France*, 2000, vol. 5133, no. I, pp. 1–15. doi: 10.1109/ASPAA.1999.810884.

[14] "Brüel & Kjær Microphone Type 4958." https://www.bksv.com/en/transducers/acoustic/microphones/special-microphones/4958

[15] "Brüel & Kjær Sound Calibrator Type 4231." https://www.bksv.com/en/transducers/acoustic/calibrators/sound-calibrator-4231

[16] Stephan Weiss, "On adaptive filtering in oversampled subbands," 1998.

[17] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 4th ed. New Jersey: Prentice-Hall, Inc, 2006. [Online]. Available: https://www.pearson.com/us/higher-education/program/Proakis-Digital-Signal-Processing-4th-Edition/PGM258227.html

[18] V. Molés-Cases, "PSZ toolbox," 2022. https://github.com/VicentMolesCases/PSZtoolbox

[19] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.*, vol. 91, no. 3, p. 1648, Aug. 1998, doi: 10.1121/1.402445.

[20] "Dynamic personal sound zones." https://www.youtube.com/watch?v=tE64tMh63NA