

Optimisation of the target sound fields for the generation of independent listening zones in a reverberant environment

Wilfried Gallian

Institute of Sound & Vibration Research
Southampton, UK
wg1v19@soton.ac.uk

Filippo Maria Fazi

Institute of Sound & Vibration Research
Southampton, UK
filippo.fazi@soton.ac.uk

Carlo Tripodi

ASK Industries S.p.A
Reggio Emilia, Italy
carlo.tripodi@askgroup.global

Nicolò Strozzi

ASK Industries S.p.A
Reggio Emilia, Italy
nicolo.strozzi@askgroup.global

Alessandro Costalunga

ASK Industries S.p.A
Reggio Emilia, Italy
alessandro.costalunga@askgroup.global

Abstract—The Pressure Matching Method (PMM) has been widely used in individual listening zones (ILZ) systems to create multiple independent sound zones in the same environment. This method aims to reproduce a set of target acoustic fields in several control zones, using one or more loudspeaker arrays. These target fields are parameters of the ILZ algorithm that are defined at the design stage. In this work, we compare the effects of several target sound field alternatives on the sound field control system performance and robustness. The proposed target fields prevent the ILZ system from focusing its effort on unessential tasks such as the dereverberation of the acoustic environment. More specifically, we compare target acoustic fields of a single active loudspeaker, of several active and in-phase loudspeakers or of a set of loudspeakers acting as a beamformer steered towards one of the zones to be controlled. Results of simulations and experiments obtained with one compact loudspeaker array in the interior of a real car are presented as well as a study on the robustness to perturbations of the solutions obtained with the various choices of target acoustic fields.

Index Terms—soundfield control, sound zone systems, pressure matching

I. INTRODUCTION

Headphones are the most popular way to enjoy our audio-visual devices in public spaces, while limiting acoustic interactions with the surrounding environment for privacy or distraction reasons. However, headphones prevent the user from communicating naturally with people around them, isolating them from the environment which can be dangerous in the case of transport. Listening through headphones may be tiring [1] or even lead to hygiene problems in the case of certain public applications. For these reasons, many research works have attempted over the past 20 years to design loudspeaker-based systems that can create individual listening zones (ILZ) inside public spaces, but without the disadvantages of headphone reproduction. In a car cabin, for

instance, the driver could use such a system to receive audio navigation instructions, with the front passenger enjoying a radio talk show, while the rear passenger could listen to music. Moreover, the information desks in an international airport could deliver the same information in different languages simultaneously to different listeners. These loudspeaker-based systems all aim to reproduce a certain audio content in a certain region of space, called the *bright zone*, while limiting the acoustic energy in another region, called the *dark zone*. Both areas can be defined either geometrically or by a discretisation of the space into control points, the pressure at which can be measured using real or virtual microphones, hereafter also referred to as control points. Two main strategies are currently used to estimate the optimal signals to be fed to these loudspeaker-based ILZ systems. The first method, called acoustic contrast control (ACC), aims to directly maximise the ratio of acoustic energy between bright and dark zone, without any consideration of the phase of the reproduced audio signal in either zone. In [2], a very slight variation considering the acoustic energy difference is presented. The second method, called pressure matching (PM), aims to reproduce given target sound pressures in both zones [3] and may be considered as a sound field synthesis technique [4]. Many examples of implementation of these two techniques can be found in the literature, in different spaces such as a car or aircraft cabin, a bank agency or simply a lounge in a family home, where there is a Hi-Fi system, a television, a desktop computer and a radio simultaneously broadcasting different content. Some works have considered multiple configurations for the loudspeaker-based systems, such as circular or linear arrays and even headrest loudspeaker systems for automotive applications [5]. Further works have also aimed at limiting the array effort of these loudspeaker arrays by means of regularisation techniques, at frequencies where the physical problem is particularly ill-conditioned [6]. Finally, compara-

tive studies have been carried out between the two methods [7], [8] with, to the best of our knowledge, only two of them discussing the selection of the target sound pressures required in the PM framework. [9] has shown that the PM method can give acoustic contrast levels close to those obtained with the ACC method, provided that appropriate target sound pressures are used. [10] compared the contrast performance obtained by the PM method for different targets produced by three different combinations of loudspeakers operating in phase. But none of these works have sought to optimise analytically the target sound pressures according to particular criteria of energy or acoustic contrast. Apart from these two attempts, unitary pulses (Dirac pulses) are chosen, in most publications, as target signals for all control points in the bright zone. This raises two main questions which will be addressed in this paper: first, why should an attempt be made to reproduce a flat frequency response at the control points in the bright zone, while being aware that reverberation exist in the acoustic space involved and that loudspeakers have physical limits in the extreme areas of the spectrum? Second, why should the system reproduce the exact same target at all control points in the bright area (i.e. same magnitude and phase) when this potentially requires a significant amount of energy from the reproduction system? In this paper, these two questions are addressed in a theoretical framework. An analytical method is presented for calculating the target sound pressures that minimise, under sound quality constraints, the effort expended by the reproduction system. Then, the properties of some usual sound pressures targets are compared with the optimal targets, in both reverberant and free-field conditions. In section II, the sound zoning problem is defined and the two main control strategies defined above are reviewed. In section III a theoretical study is presented to address the two questions above. Finally, section IV describes the results obtained both experimentally, in a reverberant environment, and theoretically, with free-field assumption.

II. GENERAL DEFINITIONS

In this section, the sound zone problem is presented along with the two main control strategies used for loudspeaker's arrays.

A. Sound zone problem definition

Fig.1 shows a specific setup with two distinct zones, A and B, located in a car cabin. The head of the driver and that of the front passenger are assumed to be located in these two zone, respectively. In practice, the general problem presented in section I can be divided into two sub-problems. The first one is when zone A is considered as a bright zone and zone B as a dark zone. As for the second one, the roles of the two zones are reversed, i.e zone B becomes the bright zone and zone A becomes the dark zone. Then, the optimal signals to be fed to the reproduction system for each of the two sub-problems can be added together to address the general problem, by linear superposition. As a consequence, in this paper, only the first sub-problem will be addressed, to avoid unnecessary

repetition. The driver's zone will then always be the bright one and the passenger's zone always the dark one. Each zone will be defined with only two control points positioned at the entrance of each ear canal of the driver and front passenger, as diagrammatically described in Fig.2.

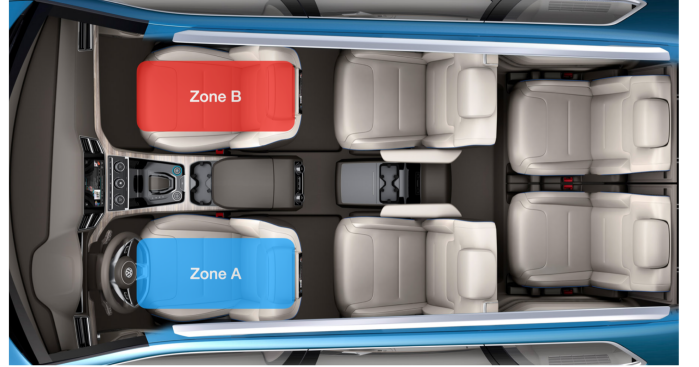


Figure 1: Car cabin setup with two distinct zones.

First, a plant matrix \mathbf{G} of transfer functions between each of the L speakers and each of the 4 control points is defined in the frequency domain, meaning that the problem is formulated and solved for one frequency at a time. However, in order to lighten the notations, any reference to the frequency index will be omitted. Thus, for any frequency, $\mathbf{G} \in \mathbb{C}^{4 \times L}$ is expressed as

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1L} \\ \vdots & \vdots & \ddots & \vdots \\ g_{41} & g_{42} & \cdots & g_{4L} \end{bmatrix} \quad (1)$$

where g_{ml} is the transfer function between the l th loudspeaker and the m th control point. Then, the following two relationships can easily be deduced as

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_b \\ \mathbf{p}_d \end{bmatrix} = \mathbf{G}\mathbf{q} = \begin{bmatrix} \mathbf{G}_b \\ \mathbf{G}_d \end{bmatrix} \cdot \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_L \end{bmatrix} \quad (2)$$

where $\mathbf{q} \in \mathbb{C}^{L \times 1}$ is the column vector of complex source strengths, $\mathbf{p} \in \mathbb{C}^{4 \times 1}$ is the column vector of the reproduced complex pressures at all control points and subscripts d and b indicate the quantities related to the dark and the bright zone, respectively.

Few evaluation metrics which will be used in the rest of this paper are defined according to [11]. First, acoustic contrast is defined as the ratio of the spatially averaged energies between the bright and dark zones and is expressed as

$$AC_{dB} = 10 \log_{10} \left(\frac{\mathbf{p}_b^H \mathbf{p}_b}{\mathbf{p}_d^H \mathbf{p}_d} \right) \quad (3)$$

where the superscript H is the conjugate transpose operator. Secondly, the array effort is expressed as

$$AE_{dB} = 10 \log_{10} \left(\frac{\mathbf{q}^H \mathbf{q}}{\mathbf{q}_{ref}^H \mathbf{q}_{ref}} \right) \quad (4)$$

where $\mathbf{q}_{ref}^H \mathbf{q}_{ref}$ is the array effort required by a single ideal monopole located at the centre of the array to reproduce the same energy as the solution \mathbf{q} in the bright area.

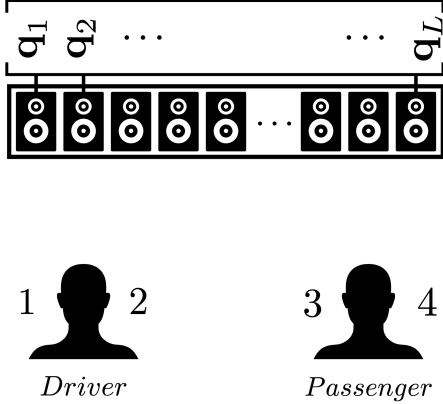


Figure 2: Diagram of a car cabin setup with a L-loudspeaker linear array in front of two listeners.

B. The Acoustic Contrast Control (ACC) method

This first technique for defining the complex source strengths aims to maximise the contrast defined in (3). For this purpose, Elliott et al. showed in [11] that a way to solve this problem while avoiding any issues caused by the ill-conditioning of the problem is to use the method of Lagrange multipliers. The objective is to minimise the energy $\mathbf{p}_d^H \mathbf{p}_d$ in the dark zone while keeping the energy in the bright zone $\mathbf{p}_b^H \mathbf{p}_b$ at a constant value E_B . Most publications also impose an energy constraint on the solution in various forms. This can be on the norm of \mathbf{q} directly or on each component of \mathbf{q} separately as in [6]. Then, the Lagrangian is expressed as

$$L = \mathbf{p}_d^H \mathbf{p}_d + \lambda_1 (\mathbf{p}_b^H \mathbf{p}_b - E_B) + \lambda_2 (\mathbf{q}^H \mathbf{q} - E) \quad (5)$$

where λ_1 and λ_2 are Lagrange multipliers and E is the array effort limit. After nullifying the partial \mathbf{q} -derivative of (5), the following eigenvalue problem arises:

$$(\mathbf{G}_d^H \mathbf{G}_d + \lambda_2)^{-1} \mathbf{G}_b^H \mathbf{G}_b \mathbf{q} = -\frac{1}{\lambda_1} \mathbf{q} \quad (6)$$

The optimal solution \mathbf{q}_{opt} is the eigenvector associated with the largest eigenvalue $-1/\lambda_1$ involved in (6). The solution must be then scaled in order to meet the energy condition on the bright zone, while not exceeding the array effort limit E . The latter can be controlled by setting the proper parameter λ_2 , which plays the same role as the Tikhonov regularisation parameter in the matrix inversion in equation (6).

C. The Pressure-Matching (PM) method

This second control method requires first of all the choice of $\mathbf{p}_T \in \mathbb{C}^{4 \times 1}$, a vector populated with the target sound pressures for all control points in both zones. For the reproduction of a single monophonic signal X_D to both ears in the bright zone and the least amount of energy possible in the dark zone, this vector takes usually the form

$$\mathbf{p}_T = \begin{bmatrix} d_1 \\ d_2 \\ 0 \\ 0 \end{bmatrix} X_D \quad (7)$$

In most papers, $d_1 = d_2 = 1$, corresponding to two Dirac pulses in the time domain. Once this choice is made, an optimisation process is run to find the complex source strengths that best replicate the target signals at all control points. In the general case, the PM method attempts to solve a conventional linear system which is

$$\mathbf{G}\mathbf{q} = \mathbf{p}_T \quad (8)$$

At this stage, it is still important to remain in the general case of a plant matrix $\mathbf{G} \in \mathbb{C}^{M \times L}$. If \mathbf{G} is fat ($M \leq L$), then the linear problem above is underdetermined i.e many choices of \mathbf{q} lead to \mathbf{p}_T . Therefore, one of the common strategies is to find the solution with minimum ℓ^2 -norm, i.e solving the following problem:

$$\min_{\mathbf{q}} \|\mathbf{q}\|_2^2 \quad s.t. \quad \mathbf{G}\mathbf{q} = \mathbf{p}_T \quad (9)$$

where $\|\cdot\|$ is the Euclidean norm. This gives the following least-norm (LN) optimal solution:

$$\mathbf{q}_{opt}^{(LN)} = \mathbf{G}^H (\mathbf{G}\mathbf{G}^H)^{-1} \mathbf{p}_T \quad (10)$$

In practice however, the $\mathbf{G}\mathbf{G}^H$ matrix can become close to singular, which causes large errors in the numerical inversion as well as impractical levels for the signals driving the speakers. Therefore, Tikhonov regularisation is often applied through diagonally loading the grammian $\mathbf{G}\mathbf{G}^H$ before inversion, with a regularisation parameter β [12]. Thus finally,

$$\mathbf{q}_{opt}^{(LN)} = \mathbf{G}^H (\mathbf{G}\mathbf{G}^H + \beta \mathbf{I})^{-1} \mathbf{p}_T \quad (11)$$

Conversely, if \mathbf{G} is skinny ($M \geq L$), the linear problem above is overdetermined and no exact solution can be reached. Therefore, one of the common strategy is to find the solution which minimises the following regularised least-squares problem:

$$\min_{\mathbf{q}} \|\mathbf{G}\mathbf{q} - \mathbf{p}_T\|_2^2 + \beta \|\mathbf{q}\|_2^2 \quad (12)$$

where β is balancing in the case the error against the size of the solution. This gives the following least-squares (LS) solution

$$\mathbf{q}_{opt}^{(LS)} = (\mathbf{G}^H \mathbf{G} + \beta \mathbf{I})^{-1} \mathbf{G}^H \mathbf{p}_T = \mathbf{G}_\beta^\dagger \mathbf{p}_T \quad (13)$$

where \mathbf{G}_β^\dagger is the regularised pseudo-inverse of \mathbf{G} . In [13], SVD analysis was applied to both underdetermined and overdetermined solutions, and in the case when regularisation is employed, it has been shown that

$$\mathbf{q}_{opt}^{(LS)} = \mathbf{q}_{opt}^{(LN)} \quad \forall \beta \quad (14)$$

Thus in the specific underdetermined problem addressed in this paper, with $L > M = 4$, the LS solution will be used by virtue of (14), for the sake of simplicity.

III. TARGET SOUND PRESSURE OPTIMIZATION

In this section, several strategies for optimising the target sound pressures defined in (7) are presented. First, the optimisation of d_1 and d_2 is presented, with the aim to spend as little energy as possible to compensate for the inherent limitations of the audio reproduction system including the loudspeaker array response and the acoustic environment. Secondly, an attempt is made to optimise the magnitude and phase relation of the two targets, trying to minimise the energy deployed by the system to reproduce both simultaneously.

A. Optimization of individual target sound pressures

In this subsection the optimisation of d_1 and d_2 is studied without analysing the effect of magnitude and phase relation between the two target pressures, as this is the subject of the next subsection. To that end, two bright zone's targets are chosen to be identical. Thus, whatever is said about the choice of d_1 may also be said about d_2 . Following this simplification, the vector of target signals can be expressed as

$$\mathbf{p}_T = \begin{bmatrix} d_1 \\ d_1 \\ 0 \\ 0 \end{bmatrix} X_D \quad (15)$$

with X_D being the input audio signal and $d_1 X_D$ the target for both driver's ears. Also, since everything is linearly dependent on X_D , this will be omitted from the calculations in the remainder of this paper.

To better understand the problem, the case of the most common target signals encountered in the literature, i.e $d_1 = d_2 = 1$, is considered. It is clear that imposing these targets at the bright zone's ears and a zero at the others requires multiple tasks from the system, namely:

- Ensuring a maximum acoustic contrast AC_{dB} between the two zones.
- Compensating for the frequency responses of the loudspeakers in the array.
- Compensating for the peaks and notches that are caused by reverberation occurring in environments such as a car cabin.

For instance, if the response of each loudspeaker approaches that of a band-pass filter, then the system will have to deploy a great amount of energy in the extreme areas of the frequency spectrum in order to avoid spectral distortion of the signals reproduced to both ears. Also, destructive and constructive interferences between the different reflected acoustic waves inside the car cabin will generate peaks and notches in the frequency response observed at the various control points. Thus, the system will have to compensate for them to reproduce a flat spectrum, by creating as many notches and peaks in the response of the signals fed to the loudspeakers. This leads to a large array effort, which in turn reduced the system's robustness to perturbations since the solution will be, in such reverberant environments, highly dependent on the control points' positions. Therefore, after the slightest

disturbance caused to the plant matrix \mathbf{G} by, for example, head's movements or an open windows, the peaks and notches will appear at difference frequencies in the transfer functions and no longer correspond to those that the system would compensate in the unperturbed scenario. As a consequence, a significant signal colouration may be heard at the control points.

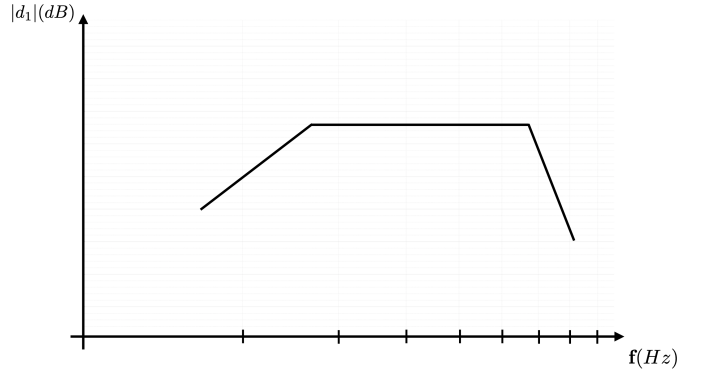


Figure 3: Magnitude response (in dB) of an ideal target taking into account the band-pass profile of a conventional loudspeaker's response.

Regarding the problem of loudspeaker limitations, the choice is made to intuitively use the response of one of the array's speakers to one of the driver's ears in the car cabin as a target signal, instead of a perfectly flat frequency response. This response, even if corrupted by the effect of the surrounding acoustics, contains the natural band-pass characteristic of each loudspeaker's response. The reproduction system will therefore no longer spend energy attempting to correct the response of the loudspeakers at the ends of the spectrum. In fact, according to (13), the speakers' complex strengths vector is expressed as

$$\mathbf{q}^{(LS)} = \mathbf{G}_\beta^\dagger \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} d_1 \quad (16)$$

This simplified form clearly shows that the ℓ^2 -norm of the solution will be directly proportional to the squared absolute value of the target d_1 . Therefore, choosing the ideal frequency response of a loudspeaker as a target, as depicted in Fig.3, instead of a perfectly flat 0 dB response, will have the effect of reducing the energy of the solution, without affecting neither the contrast AC_{dB} nor the array effort AE_{dB} . If we now seek to reduce the ℓ^2 -norm using a strong Tikhonov regularisation in frequency areas outside the power band of the loudspeakers, the AC_{dB} will also be reduced since more importance will be given to minimising the ℓ^2 -norm of \mathbf{q} than to minimising the reproduction error in (12). In conclusion, by choosing an appropriate target signal to start with, as opposed to Tikhonov regularisation strategy, it might be possible to limit the ℓ^2 -norm of the solution at the ends of the frequency spectrum without suffering any loss of AC_{dB} performance in those

regions. The results obtained with both free-field and real data will demonstrate the benefits of this method in section IV.

Secondly, regarding the problem of dereverberation of the acoustic environment, two approaches are feasible. The first is to set the same target as before, i.e. the response of one of the loudspeakers measured in the car cabin, as this will contain the main acoustic features of the car cabin such as the main resonances etc. The second is to use a frequency-dependent windowing, or trimming, of the impulse responses used to build the plant matrix \mathbf{G} , as described by Ebri et al. in [14]. This approach removes the late room reverberation before any optimisation process. However, the comparison of these two alternatives is beyond the scope of this study and represents material for future work.

B. Optimisation of magnitude and phase relation between target sound pressures

In this subsection, we return to the general case where the two bright zone's targets are different. We also introduce a filter matrix $\mathbf{W} \in \mathbb{C}^{L \times 4}$ such that

$$\mathbf{W} = \mathbf{G}_\beta^\dagger = [\mathbf{w}_1 \ \mathbf{w}_2 \ \mathbf{w}_3 \ \mathbf{w}_4] \quad (17)$$

Since this section focuses on targets only, this filter matrix will be considered unchanged for the rest of the paper. With these notations, $\mathbf{w}_1 \in \mathbb{C}^{L \times 1}$ will create a sound beam steered towards the first control point of the bright zone, being the driver's left ear, according to Fig.2. In the same manner, \mathbf{w}_2 will try to achieve the same at the driver's right ear while limiting as far as possible the acoustic energy at the other three control points. Multiplying the filter matrix \mathbf{W} by a vector \mathbf{d} populated with the target pressures allows the audio system to simultaneously reproduce two different signals at both driver's ears. Thus the optimal speakers complex strengths vector is expressed as

$$\mathbf{q} = \mathbf{W}\mathbf{d} = \mathbf{W} \begin{bmatrix} d_1 \\ d_2 \\ 0 \\ 0 \end{bmatrix} \quad (18)$$

The unnormalized array effort is defined as

$$E_q = \mathbf{q}^H \mathbf{q} = \|\mathbf{w}_1\|_2^2 |d_1|^2 + \|\mathbf{w}_2\|_2^2 |d_2|^2 + (\mathbf{w}_2^H \mathbf{w}_1) d_1 d_2^* + (\mathbf{w}_1^H \mathbf{w}_2) d_2 d_1^* \quad (19)$$

where $\|\cdot\|_2$ is defined as the Euclidean norm and a^* denotes the conjugate of $a \in \mathbb{C}$. The first two terms of (19) correspond to the energy deployed by the beamformers \mathbf{w}_1 and \mathbf{w}_2 if they were operating separately. Thus they do not depend on the phase or magnitude relationships between d_1 and d_2 . The last two terms of the sum represent the cross-energy required by the system when the two targets have to be reproduced simultaneously. After simplification,

$$E_q = \|\mathbf{w}_1\|_2^2 |d_1|^2 + \|\mathbf{w}_2\|_2^2 |d_2|^2 + 2|d_1||d_2| |\mathbf{w}_2^H \mathbf{w}_1| \cos(\Delta - \phi_w) \quad (20)$$

where

$$\Delta = \text{Arg}(d_2 d_1^*), \quad \phi_w = \text{Arg}(\mathbf{w}_2^H \mathbf{w}_1) \quad (21)$$

and where Δ is the interaural phase difference (IPD) between the two ears. The objective is now to jointly optimise $|d_1|, |d_2|$ and Δ involved in this cost function (20) in order to minimise E_q . To avoid trivial solutions, $|d_1|$ or $|d_2|$ or $|d_1|^2 + |d_2|^2$ has to be fixed during the optimisation process. In the rest of this paper, the choice was made to keep d_1 constant and the optimisation therefore corresponds to minimising the following ratio :

$$\frac{E_q}{|d_1|^2} = \|\mathbf{w}_1\|_2^2 + \|\mathbf{w}_2\|_2^2 \Gamma^2 + 2\Gamma |\mathbf{w}_2^H \mathbf{w}_1| \cos(\Delta - \phi_w) \quad (22)$$

where $\Gamma = |d_2|/|d_1|$. Note that different optimisation formulae and different results would be obtained if $|d_2|$ or $|d_1|^2 + |d_2|^2$ were chosen as a reference instead of $|d_1|$.

The objective is now to jointly optimise the two variables Δ and Γ involved in the cost function (22) in order to minimise E_q subject to a constant acoustic energy $|d_1|^2$ and without any other constraints for the moment. In (22), the normalised energy is obviously a parabolic function of the variable Γ . This parabola has a first term that is always positive by definition which implies that its vertex will be also its global minimum. For a given Δ , the vertex coordinates will then be

$$\Gamma^{(opt)} = \frac{-|\mathbf{w}_2^H \mathbf{w}_1| \cos(\Delta - \phi_w)}{\|\mathbf{w}_2\|_2^2} \quad (23)$$

and

$$\frac{E_q(\Gamma^{(opt)})}{|d_1|^2} = \left(\|\mathbf{w}_1\|_2^2 - \frac{|\mathbf{w}_2^H \mathbf{w}_1|^2 \cos^2(\Delta - \phi_w)}{\|\mathbf{w}_2\|_2^2} \right) \quad (24)$$

Yet, (23) gives a constraint on the value of Δ since a ratio Γ of modulus cannot be negative, by definition. As a consequence

$$\frac{\pi}{2} \leq (\Delta - \phi_w) \leq \frac{3\pi}{2} \quad (25)$$

As $E_q(\Gamma^{(opt)})$ will always be a positive quantity, the values of Δ that minimise (24) are

$$\Delta - \phi_w = \pi \quad \text{or} \quad \Delta - \phi_w = 0 \quad (26)$$

However, we observe that one of the two optimal solutions is not respecting condition (25) and is therefore discarded. Finally,

$$\Delta^{(opt)} = \pi + \phi_w \quad (27)$$

and

$$\Gamma^{(opt)} = \frac{|\mathbf{w}_2^H \mathbf{w}_1|}{\|\mathbf{w}_2\|_2^2} \quad (28)$$

which gives the optimal relationship between the two targets as

$$\frac{d_2^{(opt)}}{d_1^{(opt)}} = \Gamma^{(opt)} e^{j\Delta^{(opt)}} = -\frac{\mathbf{w}_2^H \mathbf{w}_1}{\|\mathbf{w}_2\|_2^2} \quad (29)$$

A color map is presented in Fig.4 for a specific frequency (2 kHz), a wide range of Δ , admissible Γ and $\phi_w = 0$, without

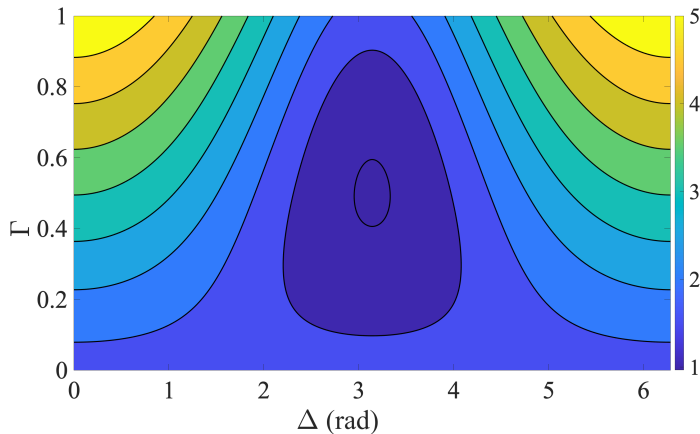


Figure 4: 2D-plot of $\frac{E_q(\Gamma^{opt})}{|d_1|^2}$ in dB.

loss of generality. This confirms the presence of a single global minimum, as predicted by the previous formula.

It should be noted that the optimal value of Δ given by equation (27) always minimises the cost function (20), whereas the optimal Γ in (28) depends on the choice of d_1 , d_2 or $|d_1|^2 + |d_2|^2$ as fixed parameter in the optimisation.

Table I: Setup distances.

Distance	value (cm)
Inter-ears distance	25
Array width	130
Shortest distance between heads and array	70
Inter-heads distance	110

IV. RESULTS

In this section, the strategies developed in the previous section are evaluated using free-field acoustic models and measurements made in a car cabin. The first two sub-sections are concerned with the evaluation of the target pressure optimisation techniques defined in section III. The third subsection is dedicated to the comparison of several usual target acoustic fields with respect to the optimal solutions.

Measurements were made in the car cabin of a standard modern car with exactly the same configuration as shown in Fig.2. Two dummy heads were facing $L = 16$ loudspeakers integrated into a compact array. This configuration is also the basis for our free-field model, which does not take into account any reflections, diffraction, head absorption or loudspeaker limitations. The main dimensions are displayed in table I, for both cases.

A. Optimisation of individual target sound pressures

In this subsection, all investigations are only made with measured data since the free-field model considers the loudspeakers as perfect source points within an infinite volume and therefore cannot be used to validate the strategies defined in

section III-A. The stake is to demonstrate the benefits in terms of E_q and AC for the reproduction system by using the target proposed in section III-A and thus without compensating for the inherent limitations of the loudspeakers. To this end, an almost non-regularised Tikhonov solution ($\beta = 10^{-5}$) was first designed using the most common target sound pressures i.e $d_1 = d_2 = 1$. For frequencies below 400 Hz, the observed array effort was very large as shown by Fig.5(b) in yellow. Beyond the pass-band behaviour of each loudspeaker in this extreme zone, another phenomenon is responsible for the increase of E_q at low frequencies. It has in fact been shown in [11] by means of a multipole expansion analysis that achieving a greater directivity requires a substantial increase in array effort, since a significant amount of destructive interferences between loudspeakers occurs at wavelengths that are larger than the size of the array. To counteract this problem, the value of the regularisation parameter β was increased at these extreme low frequencies in order to obtain an acceptable E_q profile, i.e. with the lowest possible range over the entire audible frequency spectrum. Thus, a $\beta = 10^{-1}$ was applied at low frequencies and Fig.5(b) shows in blue the E_q graph with this new β profile. In parallel to this, another solution was designed with the appropriate target defined in section III-A, i.e. close to the frequency response of one loudspeaker measured in the car cabin. Very little regularisation was used ($\beta = 10^{-5}$) in order to ensure that the appropriate targets could be perfectly reproduced. In order to facilitate a comparison between results, a small amount of equalisation was applied to this second solution in order to ensure that the two solutions had exactly the same E_q , as shown in red and blue in Fig.5(b). Fig.5(a) shows that a much higher AC score was obtained for the second solution, designed with the optimal target. The predictions reported in section III-A have thus been verified by these experimental results showing that, for a given E_q , the direct choice of an appropriate target preserves the performances in terms of AC, while Tikhonov regularisation will strongly deteriorate them.

B. Optimisation of target sound pressure ratio

In this second subsection, we first consider the special case where only Δ is considered as a variable in the cost function (20). Besides, the optimal solution Δ will be systematically compared with the usual target sound pressures ($d_1 = d_2 = 1$), corresponding to $\Delta = 0$. The two targets compared in this section will therefore initially be

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} d_1^{(opt)} \\ d_2^{(opt)} \end{bmatrix} = \begin{bmatrix} 1 \\ e^{j\Delta^{(opt)}} \end{bmatrix} \quad (30)$$

Thus, the amplitude of the targets at both ears is the same in both cases, with only the phase of one of the targets being altered. This means that the desired acoustic energy in the bright area will be identical and therefore only the numerator E_q of the AE as defined in (4) will differ between the two. Therefore, a reduction of the AE is expected because of the definition of $\Delta^{(opt)}$.

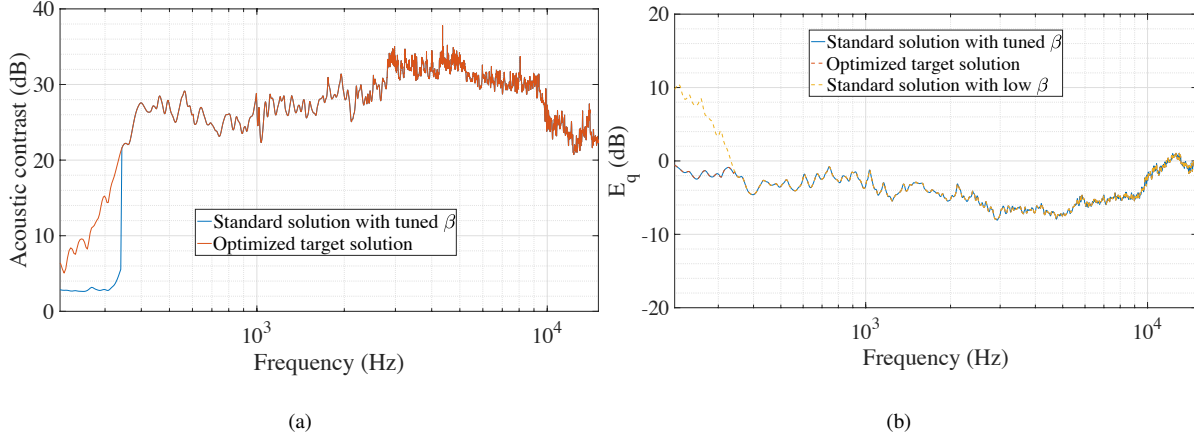


Figure 5: (a) AC (dB) (b) E_q (dB) for PM solution with optimised target signals (red line), with usual Dirac pulse targets and a frequency-dependent β (blue line), with usual Dirac pulse targets and a very low β (yellow line).

AE plots of the two different sets of targets are presented in Fig.6(a) and Fig.6(c), in both free-field and real conditions. First, low frequencies i.e below 100 Hz in free-field and 300 Hz with real data, are apparently not subject to a gain in terms of array effort. This shows that at these frequencies, a zero Δ between the pressures reproduced at the two ears is the most efficient solution. Yet, for wavelengths much greater than the head diameter, any plane wave reaching one of the two ears will show naturally little phase shift at the other ear, i.e a natural zero Δ , which proves that the optimal IPD derived is simply the natural one.

Referring now to Fig.4, it can be seen that this Δ -optimization is equivalent to moving on an horizontal line, with a static Γ . And even though it has the advantage of staying on an iso-energy line in the bright zone since $|d_1^{(opt)}|$ and $|d_2^{(opt)}|$ remain constant according to (30), the global minimum is not necessarily reached.

Therefore, the optimal parameters for Γ and Δ are now applied simultaneously, so as to reach the bluest point in Fig.4. The two targets compared in this section will therefore be

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} d_1^{(opt)} \\ d_2^{(opt)} \end{bmatrix} = \begin{bmatrix} 1 \\ \Gamma^{(opt)} e^{j\Delta^{(opt)}} \end{bmatrix} \quad (31)$$

The amplitude as well as the phase are now modified for the second control point, thus preventing the optimisation to be constrained on an iso-energy line in the bright area. Also, the desired acoustic energy in the bright area, i.e the denominator in (4), might therefore fluctuate between the two solutions, by a factor of $(1 + (\Gamma^{(opt)})^2)$ where $0 \leq \Gamma^{(opt)} \leq 1$ as observed in practice. Thus, within a few dB of accuracy, the AE will be once again similar to the unnormalised effort E_q . AE plots regarding the two different sets of targets are presented in Fig.6(b) and Fig.6(d), for both free-field and real conditions. An even greater reduction of AE is observed over the entire audible frequency spectrum, but at a potentially higher cost for the general audio quality perceived in the

bright zone, since ILD and IPD are both modified inside the optimised targets. Besides, for the same reasons as before, no additional gain is obtained in the low frequencies region, i.e below 100Hz in free-field and 200Hz with measured data. Regarding the AC performances of both solutions, Fig.7 shows a significant gain of almost 10 dB in average between 200 Hz and approximately 3 kHz, with both free-field and real data. In conclusion, selecting optimum values for both IPD and ILD will result in better acoustic contrast between the two zones for lower AE, except for low frequencies and for frequencies above 4kHz in both free-field and real cases.

Fig.8 shows a diagram of IPD versus ILD for three different solutions, for all frequencies. The first one, drawn in red, is simply the PM solution using the optimal target sound pressures. The second, drawn in yellow, is the PM solution considering only three control points, still two in the dark zone but only the first one in the bright zone, i.e with a target field reduced to $p_T = [d_1 \ 0 \ 0]$. Finally, the third, drawn in blue, is the solution obtained using the ACC method described in section II, with also only the first control point retained in the bright area. Thanks to the two equalities this figure highlights, two strong results can be inferred. First, the optimal ILD and IPD values derived in section III are identical to those observed naturally when a PM solution imposing no constraint on one of the two bright zone's control points is considered. In other words, we observe that the best target field at the driver's right ear (the second control point) is the one naturally reproduced when focusing only on the left ear. Although very intuitive, this result still requires an analytical proof, which may be the object of future work. The second result is that, as long as only one point is considered in the bright area, the PM and ACC solutions lead to the same ILD and IPD. It can in fact be shown that, subject to a condition on the energy reproduced in the bright zone, the two solutions are exactly the same. Indeed, it is sufficient to demonstrate that the three-points PM solution satisfies the eigenvalue problem that defines the three-points

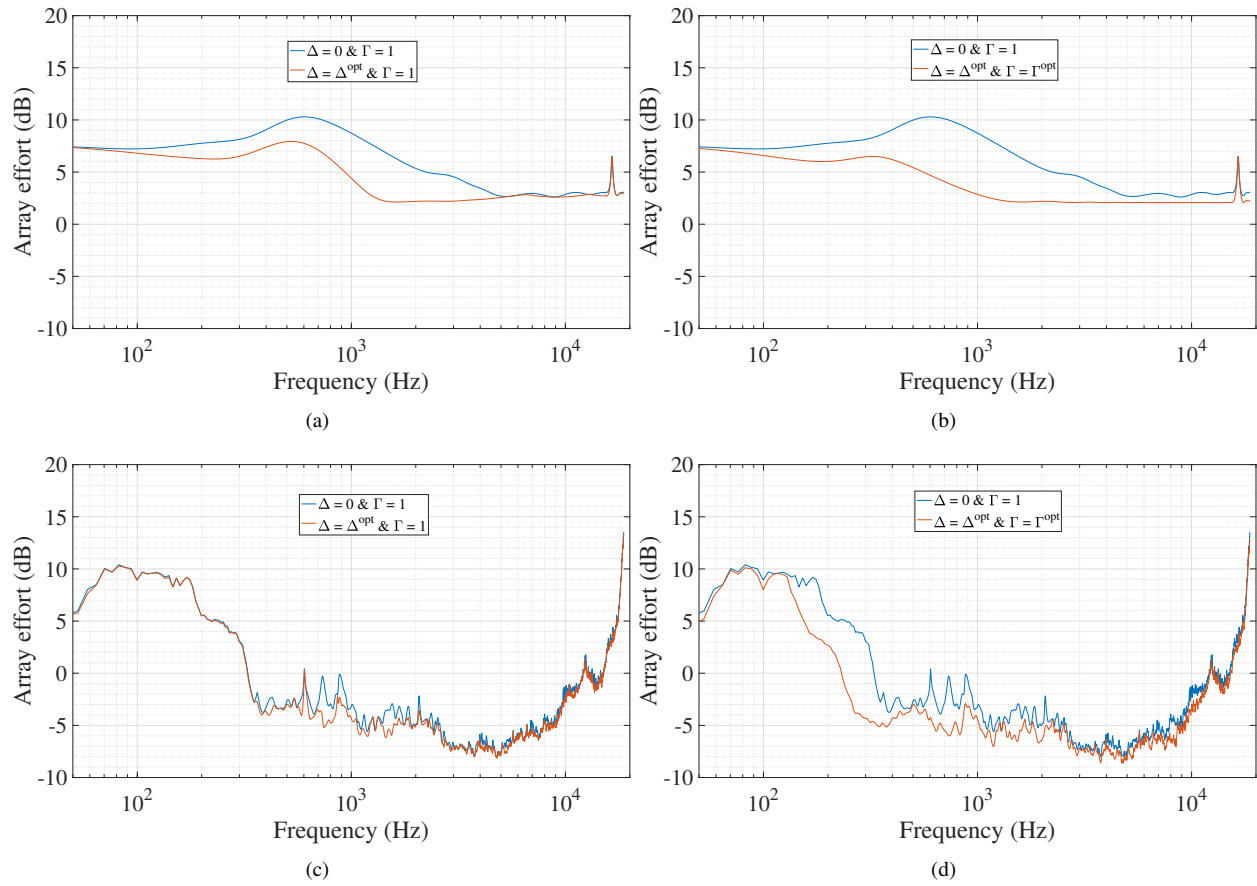


Figure 6: Array effort (AE) expressed in dB. (a) & (b) are obtained with free-field data. (c) & (d) are obtained with measured data.

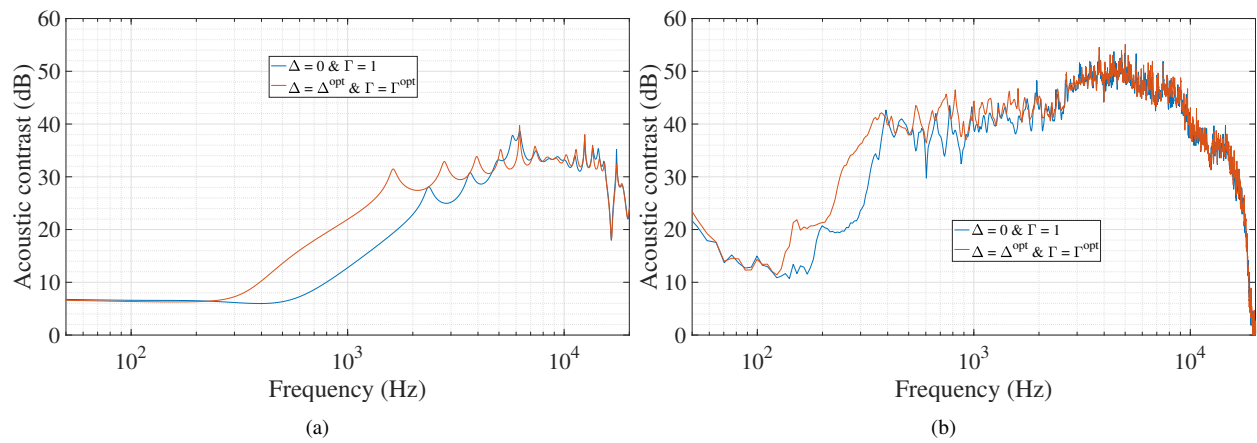


Figure 7: Acoustic contrast (AC) expressed in dB : (a) Free-field data (b) Measured data

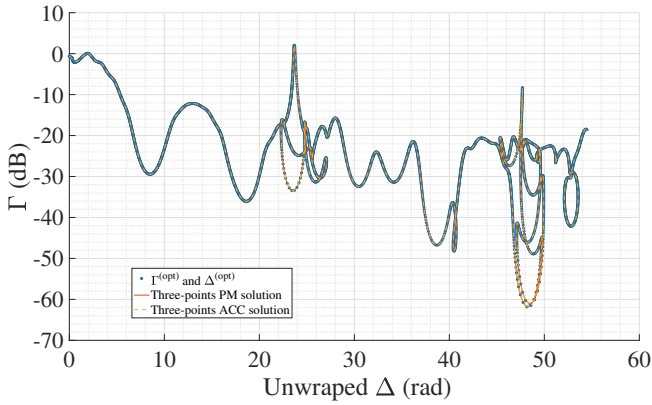


Figure 8: $\Delta - \Gamma$ relation with \mathbf{G} measured in a car cabin. In blue : PM solution with targets based on $\Delta^{(opt)}$ and $\Gamma^{(opt)}$. In yellow : ACC solution with a unique control point in the bright zone. In red : a PM solution with a unique control point in the bright zone.

ACC solution.

Yet, PM solutions are often preferred in most publications to ACC solutions for their improved phase control at each control point thus leading to better perceived audio quality, as indicated in [7]. Therefore, given the first equality shown in Fig.8, the PM solution using the optimised target will most likely lead to a degradation in the perceived audio quality similar to those observed for ACC solutions that consider only one bright control point. In practice, the system designer may want to define a solution that is psychoacoustically acceptable in terms of audio quality, but keeping the global minimum defined in Fig.4 as a reference. For this purpose, a new cost function is proposed, incorporating constraints on the admissible Γ and Δ , resulting in a limited search space in Fig.4. Using the same notations as in (20), this gives

$$\min_{\Gamma, \Delta} \frac{E_q}{|d_1|^2} \quad s.t. \quad \{\Gamma \in \Psi_\Gamma \ \& \ \Delta \in \Psi_\Delta\} \quad (32)$$

where Ψ_Γ and Ψ_Δ are the sets containing the admissible values for Γ and Δ . If these sets are infinitely narrow, there is no room to optimise the solution. This gives the usual PM solutions which may provide optimal theoretical performance in terms of audio quality, but prevents any reduction of the array effort and any consequent improvement of robustness. Conversely, if these sets are infinitely large, the freedom of movement is total, which means reaching the overall minimum while risking a significant degradation of the audio quality. The system designer can therefore choose, under real conditions and according to their own preferences, the optimal width of the two sets, in order to obtain the best trade-off between the reduction of the array effort and the fidelity to the original target chosen for the bright zone. An algorithm looking for the minimum of this new constrained optimisation problem has been implemented but experiments have yet to be carried out under real conditions with different audio quality metrics, considering different width for Ψ_Γ and Ψ_Δ .

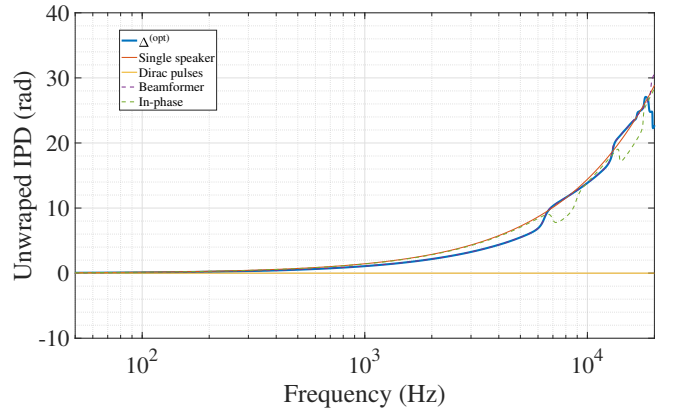


Figure 9: Unwrapped Δ for the five cases with \mathbf{G} designed in free-field.

C. Comparison of various physical target fields

In this last subsection, the IPD and ILD are compared and analysed for several target sound pressures that correspond to physical sound field obtained with one or more loudspeakers. These are then compared with the optimal IPD and ILD derived in this paper. Firstly, we consider the usual case used throughout this paper, i.e $d_1 = d_2 = 1$. Secondly, the pressure signals that are naturally reproduced by one of the array's loudspeakers, at both ears. Thirdly, the pressure signals that are reproduced by four speakers driven in phase. Finally, we investigate the case of the pressures reproduced at both ears by a beamformer steered towards the driver's left ear. Fig.10 and Fig.9 show respectively the ILD and IPD for the five target sound pressures considered. With regard to the IPD results, the beamformer is shown as being close to the optimal solution, across the whole spectrum. Indeed, the previous section demonstrated that the optimal IPD is also the one naturally reproduced by a three-point ACC solution. It is in fact to be expected that, in free field and when the dark and bright zone are far apart, the ACC will naturally create a beam of sound in the direction of the bright zone.

Secondly, the IPD reproduced by only one speaker carefully chosen in the array so as to match the blue curve, comes close to the optimal solution across the whole spectrum.

Regarding the ILD, the sound pressures reproduced by a beamformer at both ears are the only targets amongst this set that give an ILD close to the optimum. The results obtained with real data are not shown here because they led to the same general trends as described in this subsection, but their interpretation is less obvious.

V. CONCLUSION

In this work, various strategies have been presented to improve the AC and AE by optimising the target pressure signals included in the PM solution. Firstly, the method presented in section III-A prevents the system from compensating for the loudspeaker limitations, thus avoiding an explosion of the array effort at the extremities of the spectrum. To this end, the response of one of the loudspeakers was used as a

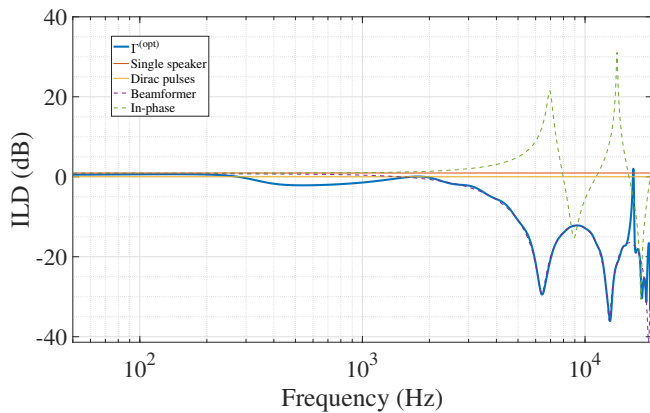


Figure 10: Γ for the five cases with \mathbf{G} designed in free-field.

target for all control points in the bright zone. Considering AE performances similar to those obtained with Tikhonov regularisation, this method proved to achieve a significant improvement to the AC, as shown in section IV. Furthermore, the relevance of dereverberation has been addressed and led to the proposal of two methods : the frequency-dependent trimming of the impulse responses and the consideration of reproduced responses by a single speaker as targets. However, no experimental validation of this aspect was carried out as part of this work and this remains an avenue to be explored in future research.

Secondly, the work has focused on the optimisation of the magnitude and phase relation (ILD and IPD) of the target sound pressures. In section IV, results obtained with optimised targets have been compared with those considering the usual targets $d_1 = d_2 = 1$. A significant decrease in AE was observed, except for the extremities of the spectrum. Furthermore, although not yet proven theoretically, results in section IV showed that the optimal target is identical to the one naturally reproduced by a PM or ACC system focusing only at one of the control points in the bright area . This indicates that the use of such optimised targets comes at a certain cost in terms of audio quality. Therefore, a new optimisation strategy that takes into account audio quality constraints has been proposed.

Finally, various physical sound fields were compared to the optimal target field. The results show that the beamformer is the system that recreates ILD and IPD in the bright zone that are closest to the optimum. A rigorous comparative study between these various soundfields involving subjective sound quality experiments may be the object of future research.

ACKNOWLEDGMENT

This work was supported by ASK Industries S.p.A. and the Italian Ministry of Economic Development (MiSE)'s fund for the sustainable growth (F.C.S.) under grant agreement (CUP) B82C21000700005, project CGS (Connettività e Guida Sicura).

REFERENCES

- [1] Thomas M Nelson and Thomy H Nilsson. "Comparing headphone and speaker effects on simulated driving". In: *Accident Analysis & Prevention* 22.6 (1990), pp. 523–529.
- [2] Mincheol Shin et al. "Maximization of acoustic energy difference between two spaces". In: *The Journal of the Acoustical Society of America* 128.1 (2010), pp. 121–131.
- [3] Ole Kirkeby and Philip A Nelson. "Reproduction of plane wave sound fields". In: *The Journal of the Acoustical Society of America* 94.5 (1993), pp. 2992–3000.
- [4] Philip Coleman. "Loudspeaker array processing for personal sound zone reproduction". PhD thesis. University of Surrey, 2014.
- [5] Stephen J Elliott and Matthew Jones. "An active headrest for personal audio". In: *The Journal of the Acoustical Society of America* 119.5 (2006), pp. 2702–2709.
- [6] PA Nelson. "A review of some inverse problems in acoustics". In: *Int. J. Acous. Vib.* 6 (2001), pp. 118–134.
- [7] Marek Olik et al. "A comparative performance study of sound zoning methods in a reflective environment". In: *Proceedings of the 52nd AES International Conference*. 2013, pp. 214–223.
- [8] Finn Jacobsen et al. "A comparison of two strategies for generating sound zones in a room". In: *18th International Congress on Sound and Vibration*. International Institute of Acoustics and Vibration. 2011.
- [9] Marcos F Simón Gálvez, Stephen J Elliott, and Jordan Cheer. "A superdirective array of phase shift sources". In: *The Journal of the Acoustical Society of America* 132.2 (2012), pp. 746–756.
- [10] Xiangning Liao et al. "Design of a loudspeaker array for personal audio in a car cabin". In: *Journal of the Audio Engineering Society* 65.3 (2017), pp. 226–238.
- [11] Stephen J Elliott et al. "Robustness and regularization of personal audio systems". In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.7 (2012), pp. 2123–2133.
- [12] Jacob Hollebon, Marcos F Simón Gálvez, and Filippo Maria Fazi. "Multiple Listener Crosstalk Cancellation Using Linear Loudspeaker Arrays For Binaural Cinematic Audio". In: *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society. 2019.
- [13] ME Gulliksson, P-Å Wedin, and Yimin Wei. "Perturbation identities for regularized Tikhonov inverses and weighted pseudoinverses". In: *BIT Numerical Mathematics* 40.3 (2000), pp. 513–523.
- [14] Michele Ebri et al. "Individual Listening Zone with Frequency-Dependent Trim of Measured Impulse Responses". In: *Audio Engineering Society Convention 149*. Audio Engineering Society. 2020.