TIME DOMAIN ACOUSTIC CONTRAST CONTROL IMPLEMENTATION OF SOUND ZONES FOR LOW-FREQUENCY INPUT SIGNALS

Daan H. M. Schellekens[1 2], Martin B. Møller[1 3], and Martin Olsen[4]

[1]Bang & Olufsen A/S, Struer, Denmark
[2]Circuits and Systems, Delft University of Technology, Delft, The Netherlands
[3]Department of Electronic Systems, Aalborg University, Aalborg, Denmark
[4]HARMAN Lifestyle Division, Struer, Denmark

d.h.m.schellekens@student.tudelft.nl, mim@bang-olufsen.dk, martin.olsen@harman.com

ABSTRACT

Sound zones are two or more regions within a listening space where listeners are provided with personal audio. Acoustic contrast control (ACC) is a sound zoning method that maximizes the average squared sound pressure in one zone constrained to constant pressure in other zones. State-of-the-art time domain broadband acoustic contrast control (BACC) methods are designed for anechoic environments. These methods are not able to realize a flat frequency response in a limited frequency range within a reverberant environment. Sound field control in a limited frequency range is a requirement to accommodate the effective working range of the loudspeakers. In this paper, a new BACC method is proposed which results in an implementation realizing a flat frequency response in the target zone. This method is applied in a bandlimited low-frequency scenario where the loudspeaker layout surrounds two controlled zones. The performance is verified with experimental results in an acoustically damped room.

Index Terms— Personal sound, sound zones, acoustic contrast control

1. INTRODUCTION

Sound zones are regions within a listening space where individual audio content is reproduced without the requirement for physical separation like wearing headphones or moving to adjacent rooms. To create such sound zones, an array of loudspeakers must be controlled, which is the topic of the presented research. In this paper, the focus is limited to a special scenario with two zones: a bright zone, where the desired sound is reproduced, and a dark zone with low sound pressure relative to the bright zone. This scenario is the simplest building block which by linear superposition can be used to create multiple spatial regions of individual audio.

In this paper, the target frequency range is restricted to low frequencies. Throughout the audible frequencies, the wave length of sound changes significantly. Due to these changes, different approaches to controlling the sound field are effective in different frequency ranges. Therefore, it can be assumed that a broadband sound zone system can be realized by combining several methods to cover the entire frequency range as proposed in [1] and further exemplified in [2].

Various methods exist to create sound zones. Two prominent examples are pressure matching (PM) [3], where both amplitude and phase of a sound field are controlled, and acoustic contrast control (ACC) [4], where the average squared sound pressure in the two zones is controlled. ACC is a sound zoning method where the average squared sound pressure is maximized in the bright zone subject to constant average squared sound pressure in the dark zone. Since only the squared sound pressure is controlled, the phase of the resulting sound field is not controlled. However, the low-frequency solution, investigated in this paper, should be part of a composite system where it is assumed that sound source localization will be dominated by the mid and high frequency solutions. Therefore, the phase of the reproduced sound field is not of concern and ACC, with the higher potential acoustical separation [5], later referred to as contrast, is the focus of this paper.

In the early publications [4, 6], ACC is defined at a single excitation frequency, resulting in solutions with a single complex weight per loudspeaker to create a bright and dark zone simultaneously. Following this approach, broadband solutions can be obtained by computing a complex weight at a set of discrete control frequencies. The inverse Fourier transform is utilized in order to create a finite impulse response (FIR) filter. In [7], Cai et al. refer to this approach as traditional acoustic contrast control (TACC). Utilizing TACC results in poor contrast performance at non-control frequencies, making a frequency domain approach ill-suited for broadband systems [7].

Solving the ACC problem for all frequencies simultaneously as an optimization formulated in the time domain is referred to as broadband acoustic contrast control (BACC) and is suggested by Elliott et al. in [8]. In [7], Cai et al. show that the BACC method results in all frequencies being filtered out, except for a single frequency, thus achieving a high acoustic contrast in the time domain. To realize a flat frequency response in the bright zone the response variation (RV) penalty term is introduced with the BACC-RV method in [7]. This formulation introduces a reference frequency which the entire frequency response should match in amplitude. The downside to this method is the requirement of determining a suitable reference frequency leading to an additional optimization. To avoid determining a reference frequency the response differentiation (RD) is introduced as a replacement for the RV-term [9]. The RD-term is a measure of the summed differences between neighboring frequency bins in the average frequency response of the resulting sound field in the bright zone.

The work in [9] considers a line array under anechoic conditions which avoids phenomena like room reflections and loudspeakers surrounding the control region. Such phenomena may cause large variation in the resulting frequency response, which in turn increases the RD-term. To avoid the RD-term being the dominating term in the optimization in a bandlimited, reverberant scenario, the term must be modified to take the expected variations into account. Reformu-

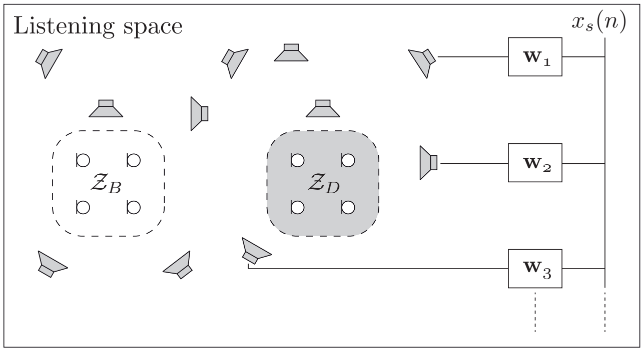

Fig. 1. Setup with $L = 11$ woofers and a square grid of 2×2 microphones per zone.

lation of the RD-term to a response trend estimation (RTE) term, resulting in a BACC-RTE method, is proposed in this paper.

The remainder of this paper is structured as follows: In Section 2, the BACC-RTE method is proposed to overcome the mentioned problems. Experimental results are presented in Section 3 for a surrounding loudspeaker layout in a damped room. The conclusion is given in Section 4.

2. THEORY

2.1. Acoustic contrast

The acoustic contrast between two zones, $C(\mathcal{Z}_B, \mathcal{Z}_D)$, is defined as the dB value of the ratio between the average squared sound pressures in the bright and dark zone [4]. Assuming zones of equal size this definition can be written as

$$C(\mathcal{Z}_B, \mathcal{Z}_D) = 10\log_{10}\left(\frac{\int_{\mathcal{Z}_B}\int |p(\mathbf{x},t)|^2 \mathrm{d}t\mathrm{d}\mathbf{x}}{\int_{\mathcal{Z}_D}\int |p(\mathbf{x},t)|^2 \mathrm{d}t\mathrm{d}\mathbf{x}}\right), \tag{1}$$

where p is the sound pressure at position $\mathbf{x}$ at time instance t and $\mathcal{Z}_B$ and $\mathcal{Z}_D$ refer to the spatial region of the bright and dark zone, respectively.

2.2. BACC-RD

A schematic of the setup is shown in Fig. 1 where L loudspeakers are driven by a source signal $x_s(n)$ filtered by a FIR filter $\mathbf{w}_l$ with length M. The room impulse response (RIR) of length I between the l-th loudspeaker ($l = 1, 2, \ldots, L$) and the k-th microphone ($k = 1, 2, \ldots, K_B$) in the bright zone is represented by $h_{B,lk}(i)$. The sampled output of the k-th microphone in the bright zone can then be expressed as

$$y_{B,k}(n) = \sum_{l=1}^{L}\sum_{i=0}^{I-1} h_{B,lk}(i) \sum_{m=0}^{M-1} w_l(m) x_s(n-m-i). \tag{2}$$

In this paper, the source signal is assumed to be spectrally white and thus only the filters $\mathbf{w}_l$ affect the input to the loudspeakers. The pressure at a microphone can therefore be described by the room impulse response from a loudspeaker to the microphone convolved by the filter $\mathbf{w}_l$ for that specific loudspeaker. This can be introduced as $x_s(n)$ being the unit sample sequence, which is assumed throughout this paper. Therefore the total impulse response of length $M+I-1$ from the source to the k-th microphone in the bright zone can be rewritten as

$$y_{B,k}(n) = \mathbf{w}^T \mathbf{r}_{B,k}(n), \tag{3}$$

where the superscript T denotes transposition, the coefficient vector $\mathbf{w}$ is defined as

$$\mathbf{w} = [w_1(0), \ldots, w_1(M-1), \ldots, w_L(0), \ldots, w_L(M-1)]^T, \tag{4}$$

and the response vector $\mathbf{r}_{B,k}(n)$ is expressed as

$$\begin{aligned}\mathbf{r}_{B,k}(n) = [&h_{B,1k}(n), \ldots, h_{B,1k}(n-M+1), \ldots \\ &\ldots, h_{B,Lk}(n), \ldots, h_{B,Lk}(n-M+1)]^T.\end{aligned} \tag{5}$$

This result can be used to compute the average squared sound pressure in the bright zone as

$$e_B = \sum_{k=1}^{K_B}\sum_{n=0}^{M+I-2} y_{B,k}^2(n)/K_B = \mathbf{w}^T \mathbf{R}_B \mathbf{w}, \tag{6}$$

where $\mathbf{R}_B$, the normalized correlation matrix in the bright zone, is expressed as

$$\mathbf{R}_B = \sum_{k=1}^{K_B}\sum_{n=0}^{M+I-2} \mathbf{r}_{B,k}(n)\mathbf{r}_{B,k}^T(n)/K_B. \tag{7}$$

Similarly, e_D and the associated $\mathbf{R}_D$ can be defined for the dark zone. To maximize the acoustic contrast in the time domain while assuming $K_B = K_D$, the BACC method [8] leads to the following contrast optimization problem

$$\begin{aligned}\mathbf{w}_{BACC} &= \arg\max_{\mathbf{w}} \frac{e_B}{e_D + \delta\mathbf{w}^T\mathbf{w}} \\ &= \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{R}_B\mathbf{w}}{\mathbf{w}^T\mathbf{R}_D\mathbf{w} + \delta\mathbf{w}^T\mathbf{w}},\end{aligned} \tag{8}$$

where $C(\mathcal{Z}_B, \mathcal{Z}_D)$ from Eqn. (1) is maximized and δ is a parameter to control the penalty on the l_2-norm of $\mathbf{w}_{BACC}$ in the optimization.

In [9] the RD-term is introduced to control the difference in sound pressure between neighboring frequency bins. Therefore, the frequency response at the k-th control point in the bright zone is defined as

$$p_{B,k}(f) = \sum_{n=0}^{M+I-2} y_{B,k}(n) e^{-j2\pi f n T_s} = \mathbf{w}^T \mathbf{s}_{B,k}(f), \tag{9}$$

where T_s is the sampling period, f a discrete frequency and $\mathbf{s}_{B,k}(f)$ is a $ML \times 1$ vector given by

$$\begin{aligned}\mathbf{s}_{B,k}(f) = [&\mathbf{r}_{B,k}(0), \ldots, \mathbf{r}_{B,k}(M+I-2)]\cdot \\ &[1, e^{-j2\pi f T_s}, \ldots, e^{-j2\pi f(M+I-2)T_s}]^T.\end{aligned} \tag{10}$$

This definition of $p_{B,k}(f)$ is then used to calculate RD, which is defined as the mean square of the first-order differential of the frequency response in the bright zone, and its formulation is given as

$$RD = \frac{1}{(J-1)K_B\Delta f}\sum_{k=1}^{K_B}\sum_{j=1}^{J-1} |p_{B,k}(f_{j+1}) - p_{B,k}(f_j)|^2, \tag{11}$$

where J is the number of discrete frequencies to be controlled and Δf denotes the frequency resolution. We suggest to pick the set of frequencies $\{f_j\}_{j=1,\ldots,J}$ as a consecutive subset of the discrete Fourier transform (DFT) frequencies, according to the bandlimited frequency range of interest. This can be done by neglecting the outer samples of the DFT frequencies. Here, the DFT frequencies are defined as an equally distributed set of $M + I - 1$ frequencies $[0, \dots, f_s - \Delta f]$, with $\Delta f = \frac{f_s}{(M+I-1)}$, where f_s is the sampling frequency.

To express RD in matrix form, let

$$\mathbf{V} = \frac{1}{\Delta f}\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{S}_k = \begin{bmatrix} \mathbf{s}_{B,k}^T(f_1) \\ \vdots \\ \mathbf{s}_{B,k}^T(f_J) \end{bmatrix}, \quad (12)$$

so that Eqn. (11) using Eqn. (9) can be rewritten as

$$RD = \mathbf{w}^T \left\{ \frac{1}{(J-1)K_B} \sum_{k=1}^{K_B} \Re\Big\{ (\mathbf{V}\mathbf{S}_k)^H (\mathbf{V}\mathbf{S}_k) \Big\} \right\} \mathbf{w}, \quad (13)$$

with $\Re\{\cdot\}$ denoting the real operator and superscript H the Hermitian transpose. When the subset $\{f_j\}_{j=1,\dots,J}$ equals the complete DFT frequency set $\mathbf{S}_k\mathbf{w}$ from Eqn. (13) is the discrete Fourier transform of $y_{B,k}(n)$. The BACC-RD filters are the solution to

$$\mathbf{w}_{RD} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{R}_B\mathbf{w}}{(1-\beta)\mathbf{w}^T\mathbf{R}_D\mathbf{w} + \beta RD + \delta \mathbf{w}^T\mathbf{w}}, \quad (14)$$

which is obtained by introducing the RD-term in Eqn. (8) and β is a weight factor between 0 and 1.

2.3. BACC-RTE

The RD-term is a function describing the accumulated changes in the resulting average frequency response in the bright zone. In a room, the transfer functions between loudspeakers and microphones will be subject to fluctuations across frequency due to reflections from room boundaries. The RD-term becomes large under such conditions and dominates the denominator in Eqn. (14). Given a limited frequency range of interest, the optimization in Eqn. (14) results in filters with high attenuation in the target frequency range and high contrast at a single frequency outside this range (as shown in Fig. 2).

To overcome this behavior, a new constraint is proposed to replace the RD-term. The response trend estimation (RTE) term is introduced to reduce the influence of local variations in the resulting frequency response. The RTE-term is a function of the difference between the means of two adjacent frequency intervals with sample length a. It can be written as

$$\begin{aligned} RTE = \frac{C_0}{\Delta f} \sum_{k=1}^{K_B} \sum_{j=1}^{J-2a+1} & |p_{B,k}(f_{j+a}) - p_{B,k}(f_j) + \dots \\ & \dots + p_{B,k}(f_{j+2a-1}) - p_{B,k}(f_{j+a-1})|^2, \end{aligned} \quad (15)$$

with $C_0 = \big((J - 2a + 1)K_B\big)^{-1}$.

Instead of $\mathbf{V}$ from the RD-term, let

$$\mathbf{V}_a = \frac{1}{a^2\Delta f}\begin{bmatrix} -1_1 & \cdots & -1_a & 1_1 & \cdots & 1_a & & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ 0 & & -1_1 & \cdots & -1_a & 1_1 & \cdots & 1_a \end{bmatrix} \quad (16)$$

be a difference matrix with dimensions $(J-2a+1) \times J$. Each entry in vector $\mathbf{V}_a\mathbf{S}_k\mathbf{w}$ is now a sum of a differences instead of a single difference between neighboring frequency bins. Each difference is calculated between two frequency bins separated by a distance of $a\Delta f$. To compensate for this distance and the number of entries in the sum $\mathbf{V}_a$ is scaled by $(a^2\Delta f)^{-1}$. Parameter a should be chosen in a such way that the smoothing interval $a\Delta f$ covers only a small local part of the frequency range of interest. Note that if $a = 1$, the RTE-term is equal to the RD-term.

The RTE-term can now be written in matrix form as

$$\begin{aligned} RTE &= \mathbf{w}^T \left\{ C_0 \sum_{k=1}^{K_B} \Re\Big\{ (\mathbf{V}_a\mathbf{S}_k)^H (\mathbf{V}_a\mathbf{S}_k) \Big\} \right\} \mathbf{w} \\ &= \mathbf{w}^T\mathbf{C}_{RTE}\mathbf{w}. \end{aligned} \quad (17)$$

This term is used to define the optimization for the BACC-RTE method by

$$\mathbf{w}_{RTE} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{R}_B\mathbf{w}}{(1-\beta)\mathbf{w}^T\mathbf{R}_D\mathbf{w} + \beta RTE + \delta \mathbf{w}^T\mathbf{w}}. \quad (18)$$

As in [10], Eqn. (18) can be rewritten as an eigenvalue problem where the solution is given as the eigenvector of

$$\big[(1-\beta)\mathbf{R}_D + \beta\mathbf{C}_{RTE} + \delta\mathbf{I}_{ML}\big]^{-1}\mathbf{R}_B, \quad (19)$$

with the largest eigenvalue.

To ensure the acoustic contrast is optimized only within the target frequency range $[f_1, f_J]$ it is recommended to filter the RIRs with a FIR bandpass filter having cut-off frequencies inside the range $[f_1, f_J]$. The filtered RIR can be used in Eqn. (5) to derive $\mathbf{R}_B$, $\mathbf{R}_D$ and $\mathbf{C}_{RTE}$.

3. EXPERIMENTAL RESULTS

3.1. Experimental setup

Fig. 1 illustrates the physical experimental setup with $L = 11$ woofers in a damped room at Bang & Olufsen (Struer, Denmark). The room is a 279 m^3 room with reverberation time below 0.6 s in the investigated frequency range. A set of RIRs from all woofers to K_B = 4 microphones in the bright and K_D = 4 microphones in dark zone was measured using sine sweeps [11]. The microphones were arranged in a square grid of 2×2 microphones with a spacing of 25 cm; approximately covering the size of the human head. The RIRs were bandpass filtered with $[f_{low}, f_{high}] = [20, 300]$ Hz, $f_s = 1.2$ kHz, and $I = 150$. BACC-RTE filters of length $M = 400$, were created with $\beta = 0.9998$, $\delta = 10^{-7}$, and $a = 2$.

The optimization of $\mathbf{w}_{RTE}$ will not give a perfectly flat frequency response in the bright zone. Therefore, equalization is introduced to create a flat frequency response in the frequency range of interest. The gain introduced with the equalizing filter is between 0 and -10 dB. The equalization violates the assumption of $x_s(n)$ being a unit sample sequence, and as a consequence the resulting contrast is suboptimal.

3.2. Results

The results presented in this section are based on two sets of RIR-measurements in the room described in Section 3.1. The first set of RIRs is used in the optimization for determining the filters. The second set of RIRs is used to evaluate the resulting performance, as presented in Fig. 2 and 3.

In Fig. 2 the gray lines show the four-microphone-average frequency responses in the bright and dark zone when applying the BACC-RTE method with the target frequency range $[f_1, f_J]$ =

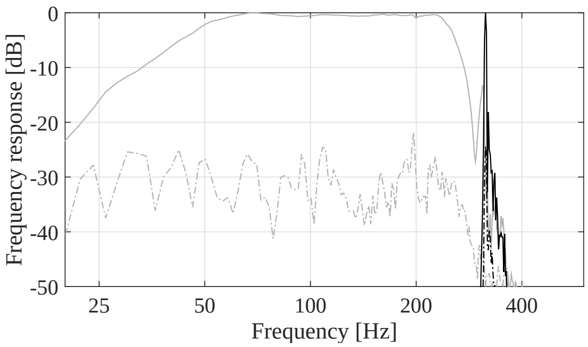

Fig. 2. Averaged frequency responses for BACC-RTE (gray) and BACC-RD (black) method in bright (—) and dark (-·-) zone with $f_1 = 10$ Hz and $f_J = 310$ Hz.

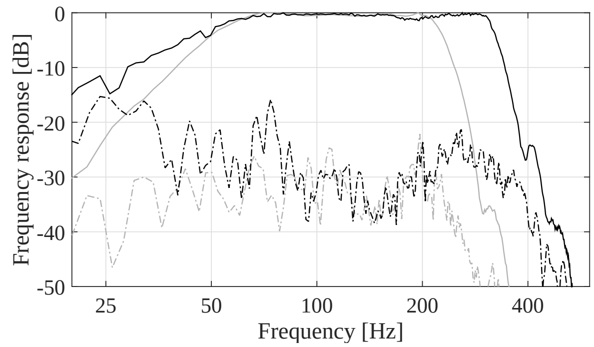

Fig. 3. Averaged frequency responses for BACC-RTE (gray) and BACC-RD (black) method in bright (—) and dark (-·-) zone with $f_1 = 0$ Hz and $f_J = f_s - \Delta f$ Hz.

$[10, 310]$ Hz, slightly wider than bandpass filter cut-off frequencies. It is observed that an acoustic contrast of almost 30 dB is achieved in the majority of the target frequency range. Furthermore, it can be seen that the contrast is optimized only within the target frequency range.

In Fig. 2 the BACC-RTE response is compared with the BACC-RD method (plotted in black). It is seen that, contrary to BACC-RTE, the BACC-RD method does not result in the desired flat frequency response, but in a solution that filters out all frequencies except a single frequency component just outside the frequency range of interest. By having almost zero sound pressure in that range the influence of the RD-term is reduced.

Frequency responses for both methods with $[f_1, f_J] = [0, \frac{f_s}{2} - \Delta f]$ are shown in Fig. 3. For the BACC-RD method β is decreased to 0.98 and the RIR length I is increased to 300 to avoid artifacts like the BACC-RD result presented in Fig. 2 due to low frequency resolution. It is seen that for both methods a flat frequency response is obtained, since it is not possible to optimize the contrast outside the frequency range $[f_1, f_J]$. However, it is generally not of interest to enforce a flat frequency response below the cut-off frequencies of the loudspeakers. To do so, would increase the risk of damaging the loudspeakers due to excessive excursion. Hereby, it is sensible to limit the constraint on the response trend to the frequency range where it is of interest to control the sound field. The BACC-RD results in Fig. 2 and 3 are very different because the BACC-RD method is not robust towards only constraining the response differences in a limited frequency range. However, the BACC-RTE results in Fig. 2 and 3 are similar, indicating that the proposed method is robust to constraining the response in a limited frequency range for the optimization. Furthermore, it can be seen from Fig. 3 that the differences between the average squared sound pressure in the bright and dark zones are similar, and in a few areas larger, for the BACC-RTE when compared to BACC-RD.

4. CONCLUSION

The aim of this work was to create a bandlimited low-frequency implementation of sound zones using time domain acoustic contrast control. It was shown that previous methods fail to realize a flat frequency response in the desired region in a bandlimited reverberant scenario. The method proposed in this paper overcomes this behavior and achieves a frequency response without sudden changes within the target frequency range.

5. REFERENCES

[1] W. F. Druyvesteyn and J. Garas, "Personal sound," *Journal of the Audio Engineering Society*, vol. 45, no. 9, pp. 685–701, 1997.

[2] J. Cheer, S. J. Elliott, and M. F. S. Gálvez, "Design and implementation of a car cabin personal audio system," *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 412–424, 2013.

[3] M. Poletti, "An investigation of 2-d multizone surround sound systems," in *Audio Engineering Society Convention 125*. Audio Engineering Society, 2008.

[4] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1695–1700, 2002.

[5] P. Coleman, P. J. Jackson, M. Olik, M. Olsen, M. B. Møller, J. A. Pedersen, et al., "The influence of regularization on anechoic performance and robustness of sound zone methods," in *Proceedings of Meetings on Acoustics*. Acoustical Society of America, 2013, vol. 19, p. 055055.

[6] M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. H. Park, and J. Seo, "Maximization of acoustic energy difference between two spaces," *The Journal of the Acoustical Society of America*, vol. 128, no. 1, pp. 121–131, 2010.

[7] Y. Cai, M. Wu, and J. Yang, "Design of a time-domain acoustic contrast control for broadband input signals in personal audio systems," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 341–345.

[8] S. J. Elliott and J. Cheer, "Regularisation and robustness of personal audio systems," ISVR Technical Memorandum 995, University of Southampton, 2011.

[9] Y. Cai, M. Wu, L. Liu, and J. Yang, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. EL252–EL257, 2014.

[10] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 2123–2133, 2012.

[11] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*. Audio Engineering Society, 2007.