DESIGN OF A TIME-DOMAIN ACOUSTIC CONTRAST CONTROL FOR BROADBAND INPUT SIGNALS IN PERSONAL AUDIO SYSTEMS

Yefeng Cai, Ming Wu, Jun Yang

Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

ABSTRACT

The acoustic contrast control (ACC) approach provides a simple strategy to focus sound on a designed target area in a personal audio system. All of the traditional ACC approaches have been designed in the frequency domain, and then transformed into the time domain through the application of an inverse Fourier transform. Therefore, the broadband contrast performance of the traditional ACC approach is not optimum. Especially, when the length of the control filter is short, the traditional approach may have poor contrast performance at non-control frequencies. This study proposes a novel method that can achieve the optimum broadband contrast performance, which also maintains a good frequency-response consistency by introducing the response variation constraints to improve sound quality. Experimental results demonstrate the effectiveness of the new method.

Index Terms— Acoustic contrast control, Frequency response consistency, Response variation, Loudspeaker array, Personal audio systems

1. INTRODUCTION

Personal audio systems can focus sound on a bright zone and provide a private listening space for users without disturbing others in an adjacent dark zone [1]. Several control schemes can be applied to this system, such as multi-zone reproduction techniques [2–3], beamforming [4-5], and acoustic contrast control (ACC) [1, 6–9]. The ACC approach can effectively provide an optimal solution that maximizes acoustic contrast, which represents the acoustic energy ratio between bright and dark zones. Chang et al. [6] verified that the ACC approach is feasible. Furthermore, Park et al. [7] demonstrated that it is applicable in a two-channel private audio system. Shin et al. [8] proposed a similar method that enables the maximization of the acoustic energy difference between two selected zones to eliminate the ill-conditioning problem in the ACC approach. Elliott et al. [9] investigated the sensitivity of the ACC approach to changes in the acoustic environment and improved its robustness by using a regularization method.

Traditional acoustic contrast control (TACC) approaches are computed at a complete set of discrete control frequencies. The impulse response of the filter is then obtained by an inverse Fourier transform of the frequency response. The resulting design cannot guarantee the causality of the time-domain filter. In addition, the maximization of acoustic contrast at non-control frequencies and the consistency of the frequency-response also cannot be guaranteed, especially if the length of filter is short. Thus, the TACC approach may not be suitable for broadband systems. To design a broadband acoustic contrast control (BACC) over a continuous frequency range, Elliott et al. [10] suggested that TACC should be converted to a time-domain digital-filter problem. A potential problem for the BACC approach is that the frequency response consistency still can not be controlled, which may lead to the intolerable distortion and degradation of sound quality.

This paper presents a novel method that addresses both the problems of controlling the acoustic contrast and maintaining a good frequency-response consistency in the bright zone for broadband input signals. The novel method introduces a response variation (RV) [11] term into the BACC approach design to solve the frequency-response consistency problem and improve the sound quality in the bright zone. In the succeeding sections of this paper, the novel BACC-RV approach is described in detail. Performance evaluation of this approach is performed based on experimental results and compared with TACC and BACC approaches.

2. PROBLEM STATEMENT

The BACC approach is used to adjust the amplitude and the phase of the broadband signal that drives each loudspeaker with a finite impulse response (FIR) filter of length M, as shown in Fig. 1. The acoustic contrast between bright and dark zones in a continuous spectrum is then maximized using an array of L loudspeakers. The bright zone is represented by K discrete points. The impulse response between the lth loudspeaker element and the kth control point in the bright zone is modeled as an FIR filter of length I. The ith coefficient of the impulse response is written as $h_{\mathrm{B}lk}(i)$. The sampled output $y_{\mathrm{B}k}(n)$ at the kth control point

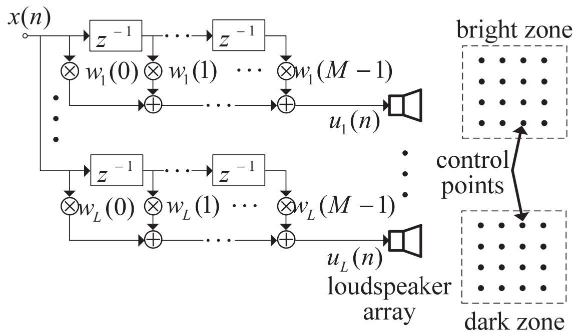


Fig. 1. General structure of the broadband acoustic contrast control method

can be expressed as follows:

$$y_{Bk}(n)=\sum_{l=1}^{L}\sum_{i=0}^{I-1}h_{Blk}(i)u_l(n-i) \tag{1}$$

$$u_l(n)=\sum_{m=0}^{M-1}w_l(m)x(n-m) \tag{2}$$

where $w_l(m)$ is the mth coefficient of the FIR filter for the lth loudspeaker; $u_l(n)$ and $x(n)$ are the input signals for the lth loudspeaker and the loudspeaker array system, respectively.

We can substitute (2) into (1). Thus, $y_{Bk}(n)$ can be expressed as follows:

$$y_{Bk}(n)=\sum_{l=1}^{L}\sum_{m=0}^{M-1}w_l(m)[\sum_{i=0}^{I-1}h_{Blk}(i)x(n-m-i)] \tag{3}$$

(3) can be expressed more concisely as follows:

$$y_{Bk}(n)=\mathbf{w}^T\mathbf{r}_{Bk}(n) \tag{4}$$

where $\mathbf{w}$ is the coefficient vector given by

$$\mathbf{w}=[w_1(0),\cdots,w_1(M-1),\cdots,w_L(0),\cdots,w_L(M-1)]^T \tag{5}$$

and the filtered signal vector $\mathbf{r}_{Bk}$ is defined as follows:

$$\begin{aligned}\mathbf{r}_{Bk}(n)=[r_{B1k}(n),\cdots r_{B1k}(n-M+1),\cdots,r_{BLk}(n),\\ \cdots r_{BLk}(n-M+1)]^T\\ r_{Blk}(n)=\sum_{i=0}^{I-1}h_{Blk}(i)x(n-i)\end{aligned} \tag{6}$$

To design the BACC, the input signal $x(n)$ is considered as white noise. Thus, the average acoustic energy e_B in the bright zone can be written as follows:

$$e_B=\frac{1}{K}\sum_{k=1}^{K}E\{y_{Bk}(n)y_{Bk}(n)\}=\mathbf{w}^TE\{\mathbf{R}_B^T(n)\mathbf{R}_B(n)\}\mathbf{w} \tag{7}$$

where $E\{\bullet\}$ denotes the expectation operator. The normalized filtered signal matrix $\mathbf{R}_B(n)$ is expressed as:

$$\mathbf{R}_B(n)=\frac{1}{\sqrt{K}}\begin{bmatrix}\mathbf{r}_{B1}^T\\ \vdots\\ \mathbf{r}_{BK}^T\end{bmatrix} \tag{8}$$

Similarly, the averaged acoustic energy e_D in the dark zone can be written as follows:

$$e_D=\mathbf{w}^TE\{\mathbf{R}_D^T(n)\mathbf{R}_D(n)\}\mathbf{w} \tag{9}$$

where $\mathbf{R}_D(n)$ is the normalized filtered signal matrix in the dark zone. The acoustic contrast in the time domain can be defined as

$$C_t=\frac{e_B}{e_D}=\frac{\mathbf{w}^TE\{\mathbf{R}_B^T(n)\mathbf{R}_B(n)\}\mathbf{w}}{\mathbf{w}^TE\{\mathbf{R}_D^T(n)\mathbf{R}_D(n)\}\mathbf{w}} \tag{10}$$

Therefore, the BACC design leads to the following optimization problem:

$$\max_{\mathbf{w}}\quad\frac{\mathbf{w}^TE\{\mathbf{R}_B^T(n)\mathbf{R}_B(n)\}\mathbf{w}}{\mathbf{w}^TE\{\mathbf{R}_D^T(n)\mathbf{R}_D(n)\}\mathbf{w}+\delta\mathbf{w}^T\mathbf{w}} \tag{11}$$

where the regularization term $\mathbf{w}^T\mathbf{w}$ is added to improve the robustness and δ is a preselected regularization parameter. The optimal solution of (11) can be obtained by finding the eigenvector that corresponds to the maximum eigenvalue of the matrix $[E\left\{\mathbf{R}_D^T(n)\mathbf{R}_D(n)\right\}+\delta\mathbf{U}]^{-1}E\left\{\mathbf{R}_B^T(n)\mathbf{R}_B(n)\right\}$ [9], where $\mathbf{U}$ is the identity matrix.

However, the designs of BACC approach do not consider the frequency-response consistency, thereby causing the sound quality to degrade in the bright zone.

3. BACC-RV

To address the problem in the BACC approach, an RV term that measures the fluctuation of the frequency response is introduced into the proposed BACC-RV approach. The RV term controls the specific frequency-response consistency in the bright zone. The frequency response $p_{Bk}(f)$ as a function of frequency f at the kth control point is defined as follows:

$$p_{Bk}(f)=\sum_{l=1}^{L}[\sum_{m=0}^{M-1}w_l(m)e^{-j2\pi fmT_s}]g_{Blk}(f) \tag{12}$$

where T_s is the sampling period and $g_{Blk}(f)$ denotes the transfer function between the lth loudspeaker element and the kth control point in the bright zone. The frequency response can be simplified as in a vector form

$$p_{Bk}(f)=\mathbf{w}^T\mathbf{s}_{Bk}(f)\ , \tag{13}$$

where $\mathbf{s}_{Bk}(f)$ is the $ML\times1$ vector as expressed in the following equation:

$$\mathbf{s}_{\mathrm{B}k} = [g_{B1k}(f), \cdots, g_{B1k}(f) e^{-j2\pi f (M-1)T_s}, \cdots g_{BLk}(f), \\ \cdots, g_{BLk}(f) e^{-j2\pi f (M-1)T_s}]^T \tag{14}$$

Consequently, RV is defined as follows:

$$\begin{aligned} RV &= \frac{1}{K} \frac{1}{B_\Omega} \sum_{k=1}^{K} \sum_{f \in \Omega} |\mathbf{w}^T \mathbf{s}_{\mathrm{B}k}(f) - \mathbf{w}^T \mathbf{s}_{\mathrm{B}k}(f_r)|^2 \\ &= \mathbf{w}^T \Re\{\mathbf{Q}^H \mathbf{Q}\} \mathbf{w} \end{aligned} \tag{15}$$

where $\Re\{\bullet\}$ stands for the real part operator , Ω is the frequency range of interest, B_Ω denotes the number of uniformly sampled frequency points over Ω , and f_r is the reference frequency. $\mathbf{Q}$ is defined as follows:

$$\mathbf{Q} = \frac{1}{\sqrt{KB_\Omega}} \begin{pmatrix} \mathbf{s}_{\mathbf{B}1}^T(f) - \mathbf{s}_{\mathbf{B}1}^T(f_r) \\ \vdots \\ \mathbf{s}_{\mathbf{B}K}^T(f) - \mathbf{s}_{\mathbf{B}K}^T(f_r) \end{pmatrix} \tag{16}$$

RV represents the frequency response variation in the bright zone and should be as small as possible. Thus, the proposed BACC-RV is expressed as follows:

$$\max_{\mathbf{w}} \quad \frac{\mathbf{w}^T E\{\mathbf{R}_{\mathrm{B}}^T(n) \mathbf{R}_{\mathrm{B}}(n)\} \mathbf{w}}{\mathbf{w}^T E\{\mathbf{R}_{\mathrm{D}}^T(n) \mathbf{R}_{\mathrm{D}}(n)\} \mathbf{w} + \delta \mathbf{w}^T \mathbf{w} + \beta RV} \tag{17}$$

where β is a weight factor that creates a trade-off between the frequency-response consistency and the acoustic contrast in the time domain.

Similarly to (11), the optimal solution of (17) is the eigenvector that corresponds to the largest eigenvalue of the matrix

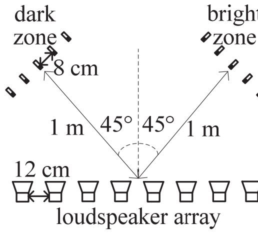
$$[E\left\{\mathbf{R}_{\mathrm{D}}^T(n) \mathbf{R}_{\mathrm{D}}(n)\right\} + \delta \mathbf{U} + \beta \Re\{\mathbf{Q}^H \mathbf{Q}\}]^{-1} E\left\{\mathbf{R}_{\mathrm{B}}^T(n) \mathbf{R}_{\mathrm{B}}(n)\right\}.$$

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Experimental setup

The experiment is conducted in an anechoic chamber (Fig. 2) at the Institute of Acoustics, Chinese Academy of Sciences (China). The linear loudspeaker array consists of eight moving coil speaker units that are equally spaced at 12 cm. The bright zone and the dark zone are located in the 45° and −45° deviation direction from the center, respectively. Each zone is defined by a five-element microphone array at a spacing of 8 cm. The loudspeaker and the microphone arrays are positioned on the same horizontal plane with a center spacing of 1 m.

This paper only considers the processing of the speech signal to set the system sampling rate at 8 kHz. The time length of the impulse response is determined to be 0.1 s, which is enough for the free-field condition. Thus the length I of impulse response is set to 800. The impulse response and transfer function between the loudspeaker and microphone arrays are initially measured. The coefficient vector is then computed according to TACC, BACC, and BACC-RV approaches. The performances of the acoustic contrast at each frequency and frequency-response consistency are compared based on the obtained impulse responses and transfer functions.



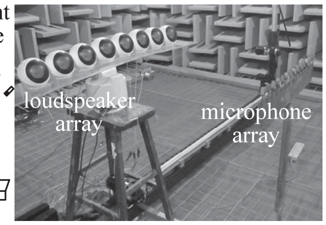


Fig. 2. (a) Schematic diagram of the experimental setup, (b) experimental setup in the anechoic chamber

Frequency-response consistency can also affect the sound quality in the bright zone. Sound quality is determined based on the perceptual evaluation of speech quality (PESQ) [12] values. Two English sentences, “the birch canoe slid on the smooth planks” and “glue the sheet to the dark blue background” (e1 and e2, respectively), are delivered by a male speaker and used as original source signals.

The regularization parameter δ and weight factor β are set at 0.01 and 0.05, respectively. The reference frequency is 1 kHz. The length M of the FIR filter is set to 50 or 200, and the control frequency interval for the TACC method is 160 or 40 Hz, respectively.

4.2. Results and discussion

The acoustic contrast at each frequency with three different methods is shown in Fig. 3 (The comparison performance of the acoustic contrast is plotted separately in order to see more clearly). To demonstrate the acoustic contrast performance at non-control frequencies for the TACC method, the investigated frequency interval in Fig. 3 is 10 Hz. The performance of the TACC method is the worst among the three methods. Although the acoustic contrast at the control frequency is high, the acoustic contrast rapidly decreases when the frequency deviates from the control frequency point. The BACC method partially alleviates this problem. Compared with the other methods, the proposed BACC-RV method performs best in almost all of the frequencies except the control frequencies.

The frequency response at the center of the microphone array in the bright zone is illustrated in Fig. 4. The fluctuation of the frequency response with the TACC method is large over the entire frequency range. This is because the TACC method does not consider the frequency-response consistency in the bright zone. The performance of the frequency response with the BACC method is the worst

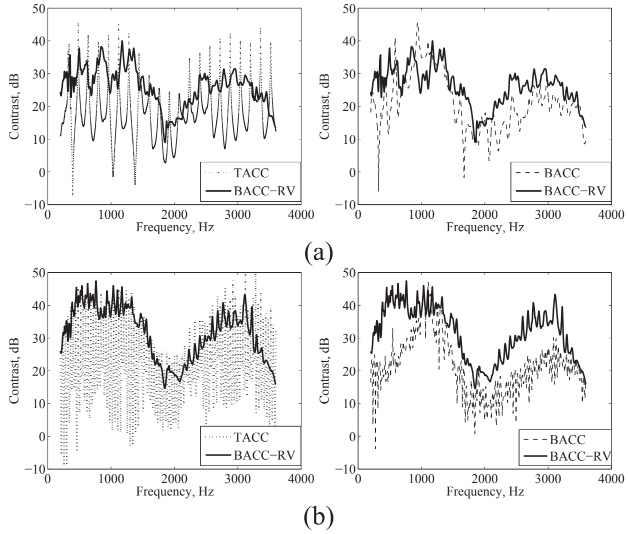
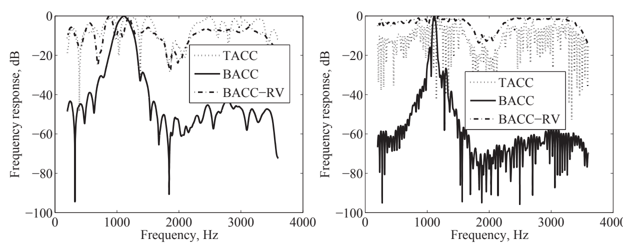


Fig. 3. Acoustic contrast versus frequency with traditional acoustic contrast control (TACC), broadband acoustic contrast control (BACC), and BACC-response variation (RV) methods. Length M of the finite impulse response (FIR) filter: (a) 50 and (b) 200.



(a) (b)

Fig. 4. Frequency response at the center of the microphone array in the bright zone with traditional acoustic contrast control (TACC), broadband acoustic contrast control (BACC), and BACC-response variation (RV) methods. Length M of the finite impulse response (FIR) filter: (a) 50 and (b) 200.

among the methods. It can be observed that only the frequency component around the 1 kHz is reserved. The BACC method considers the maximization of the acoustic contrast in the time domain. As a result, the BACC method only retains the frequency with the highest acoustic contrast observed in the frequency domain. Therefore, the frequency response is more similar to line spectrum as the length M of the FIR filter increases (Fig. 4), which results in poor frequency-response consistency. The BACC-RV method can alleviate this problem with the help of the RV constraint to achieve the best frequency-response consistency compared with the other two methods.

Table 1 shows the PESQ values between the source speech signals and the speech signals achieved at the center of the microphone array in the bright zone. Compared with the other two methods, the BACC-RV method produces the highest PESQ values and improves sound quality significantly.

Table 1. Perceptual evaluation of speech quality (PESQ) values between the source speech signals and speech signals achieved at the center of the microphone array in the bright zone.

		TACC	BACC	BACC-RV
M = 200	e1	3.040	2.907	4.270
	e2	3.030	2.824	4.305
M = 50	e1	3.796	3.807	4.188
	e2	3.791	3.719	4.290

5. CONCLUSION

This study focuses on the formulation of the ACC algorithm, which can generate a focused sound field in a personal audio system by using a loudspeaker array. Most of previous studies [1, 6–9] have only considered the acoustic contrast within a series of discrete control frequencies. Recently, Elliott et al. [10] defined the acoustic contrast in the time domain and described a time-domain design, which corresponds to the BACC approach in the current study. Although no experiment or simulation results were provided in the previous report, the current paper shows that the BACC method can partially alleviate the contrast problem at non-control frequencies. The novel BACC-RV method, which is related to the BACC approach, is directly designed in the time domain and can control the contrast within a continuous frequency range. This study also introduces an RV term to control the frequency-response consistency and prevent the severe degradation of sound quality, which was not considered in previous studies.

Considering a proper selection of δ and β, the proposed BACC-RV method can efficiently make the acoustic contrast better over a continuous frequency range. A good frequency-response consistency can also be produced to improve the sound quality in the bright zone compared with TACC and BACC methods.

6. ACKNOWLEDGMENTS

This work was supported by National Natural Science Fund of China under Grants No.11004217, 11074279, 11174317.

7. REFERENCES

[1] 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, no. 4, pp. 1695-1700, 2002.

[2] M. Poletti, "An investigation of 2D multizone surround sound systems," in *Proc. 125th Conv. AES*, San Francisco, CA, Oct. 2008.

[3] T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *Proc. 2011 IEEE Int.*

Conf. Acoust. Speech Signal Process. (*ICASSP)*, May, 2011, pp. 437-440.

[4] Y. Hur, S. W. Kim, Y. Park and D. H. Youn, "Highly focused sound beamforming algorithm using loudspeaker array system," in *Proc. 125th Conv. AES*, San Francisco, CA, Oct. 2008.

[5] I. Tashev, J. Droppo, M. Seltzer, and A. Acero, "Robust design of wideband loudspeaker arrays," in *Proc. 2008 IEEE Int. Conf. Acoust. Speech Signal Process.* (*ICASSP)*, April, 2008, pp. 381-384.

[6] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, "A realization of sound focused personal audio system using acoustic contrast control," *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2091-2097, 2009.

[7] J.-Y. Park, J.-H. Chang, and Y.-H. Kim, "Generation of independent bright zones for a two-channel private audio system," *J. Audio Eng. Soc.*, vol. 58, no. 5, pp. 382-393, 2010.

[8] M.-C. Shin, S.-Q. Lee, F. M. Fazi, P. A. Nelson, D.-S. Kim, S.-M. Wang, K.-H. Park, and J.-I. Seo, "Maximization of acoustic energy difference between two spaces," *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 121-131, 2010.

[9] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 7, pp. 2123-2133, 2012.

[10] S. J. Elliott and J. Cheer, "Regularization and robustness of personal audio systems," *ISVR Technical Memorandum 995*, [Online]. Available: http://eprints.soton.ac.uk/207989/1/Pub12685.pdf, (Accessed 3 November, 2012).

[11] Y. Zhao and W. Liu, "Robust wideband beamforming with frequency response variation constraint subject to arbitrary norm-bounded error," *IEEE Trans. Antennas Propag.*, vol. 60, no. 5, pp. 2566-2571, 2012.

[12] ITU P.862. Perceptual evaluation of speech quality (PESQ) and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. ITU-T Recommendation P.862, 2000.