24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ICSV24
23-27JULY 2017
LONDON CALLING

SOUND ZONES: ON ENVELOPE SHAPING OF FIR FILTERS

Martin Bo Møller
Aalborg University, Department of Electronic Systems, Aalborg, DK
Bang & Olufsen, Struer, DK
email: marbm@es.aau.dk

Martin Olsen
Harman Lifestyle Audio, Struer, DK

Sound zones, i.e. spatially confined regions of individual audio content, can be created by appropriate filtering of the desired audio signals reproduced by an array of loudspeakers. The challenge of designing filters for sound zones is twofold: First, the filtered responses should generate an acoustic separation between the control regions. Secondly, the pre- and post-ringing as well as spectral deterioration introduced by the filters should be minimized. The tradeoff between acoustic separation and filter ringing is the focus of this paper. A weighted L2-norm penalty is introduced in the sound zones optimization problem, to reduce pre- and post-ringing of the filters. The effect of shaping the filter envelopes is investigated in an experimental setup with eight woofers, surrounding two control regions. The results show that it is possible to reduce the pre- and post-ringing of the filters without significantly reducing the acoustic separation between the zones.

Keywords: sound zones, personal audio, sound field control.

1. Introduction

The concept of sound zones concerns the scenario where individual audio content is provided to multiple listeners in a given listening space, without using room dividers or headphones. This can be achieved by controlling an array of loudspeakers in order to generate the desired multi zone sound field. The control scheme often utilized is based on defining a set of finite impulse response (FIR) control filters, specifically applied for each individual loudspeaker. To provide individual audio content in two or more separate spatial regions, one zone is defined as the target, or equivalently the bright zone, while all other zones are referred to as dark zones. The dark zone(s) has low average squared sound pressure, relative to the bright zone. The creation of sound zones with different audio content, can then be realized by the principle of superposition. Due to the variation in wave length throughout the audible frequencies, different control strategies must be combined and applied as a composite solution to cover the entire audible frequency range, as suggested in [1]. The scope of the current work is limited to the creation of sound zones at low frequencies (below 300 Hz), to comply with the experimental setup of eight 10" woofers surrounding two zones.

The primary performance parameter for the evaluation of sound zones is the contrast [2], i.e. the ratio of average squared sound pressure in the bright zone relative to the dark zone. In the perceptual evaluations of the sound zoning performance, a parameter has been introduced as target-to-interferer ratio (TIR) which has been related to the attributes annoyance and distraction [3, 4]. This introduces the evaluation of the interference, whereas the evaluation of the target audio has been limited to the mean square pressure in the bright zone and the mean square error between a desired sound field and

the reproduced sound field [5]. Recently, the sound zone performance metrics has been extended by speech privacy and intelligibility [6].

In a recent publication [7], multiple methods applied for generating low-frequency sound zones were compared. The outcome was that comparable acoustic separation performance could be attained with all the investigated methods. One difference between the control methods was the envelope of the control filters in the time-domain. The assumption in this paper is that the envelope of the control filters is related to the perceived audio quality of the target sound field in the bright zone. For instance, if the envelope of the filters approaches a square window, significant temporal artifacts are expected to occur while an envelope resembling the delta function is likely to be perceptually neutral. In relation to this, it is of interest to compare the obtainable contrast relative to the specific envelope penalty applied to the control filters.

2. Theory

The common metric for evaluating the sound field separation between the bright and dark zone is the acoustic contrast defined in the frequency domain by

$$\mathrm{Contrast}(\omega) = 10\log_{10}\left(\frac{K_D\sum_{K_B}|p_k(\omega)|^2}{K_B\sum_{K_D}|p_k(\omega)|^2}\right), \tag{1}$$

where $p_k(\omega)$ is the complex sound pressure at the k^{th} spatial position at angular frequency ω. In the following it is assumed that the number of sampling points in each zone is equal ($K_B = K_D$), hence the normalization is omitted.

When the design of control filters is included in the sound zoning problem definition, it is beneficial to consider a time-domain formulation [7]. The data structure, introduced here, relates the control filters, loudspeakers, and control points in the listening room. The applied control scheme assumes a linear time-invariant system, which can be controlled with a feed-forward system. The impulse response at a single microphone position, due to the sum of filtered loudspeaker responses, can be expressed in matrix notation as

$$\mathbf{p}_k = \mathbf{H}_k\mathbf{w}, \tag{2}$$

where $\mathbf{H}_k$ is a block matrix consisting of convolution matrices describing the impulse response from each loudspeaker to the observation point, and can be written as

$$\mathbf{H}_k = \begin{bmatrix}\mathbf{H}_{k,1} & \cdots & \mathbf{H}_{k,L}\end{bmatrix} \tag{3}$$

$$\mathbf{H}_{k,l} = \begin{bmatrix} h_{k,l}(0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ h_{k,l}(I-1) & \ddots & h_{k,l}(0) \\ 0 & \ddots & \vdots \\ 0 & 0 & h_{k,l}(I-1)\end{bmatrix}. \tag{4}$$

The FIR filters of length M are collected as a concatenated vector in the form

$$\mathbf{w} = \begin{bmatrix}\mathbf{w}_1^T & \cdots & \mathbf{w}_L^T\end{bmatrix}^T \tag{5}$$

$$\mathbf{w}_l = \begin{bmatrix}w_l(0) & \cdots & w_l(M-1)\end{bmatrix}^T. \tag{6}$$

The concatenated pressure impulse responses for all the points sampling the bright zone, can then be expressed as

$$\mathbf{p}_B = \mathbf{H}_B\mathbf{w} \tag{7}$$

$$\mathbf{H}_B = \begin{bmatrix}\mathbf{H}_1^T & \cdots & \mathbf{H}_{K_B}^T\end{bmatrix}^T. \tag{8}$$

Hereby, the total squared energy in the bright zone can be described as the inner product given by

$$\mathbf{p}_B^T\mathbf{p}_B = \mathbf{w}^T\mathbf{R}_B\mathbf{w} = \mathbf{w}^T\mathbf{H}_B^T\mathbf{H}_B\mathbf{w}. \tag{9}$$

Here, $\mathbf{R}_B$ represents the sum of cross-correlations between the loudspeaker responses at each observation point in the bright zone. The corresponding matrix representing the cross-correlations at observation points in the dark zone is introduced as $\mathbf{R}_D$. Thus, expressions for the energy in the bright and dark zone have been established, which can be used to formulate an optimization problem in terms of the energy distribution in the zones. A constraint is introduced in order to ensure the solution controls the entire frequency range of interest, as addressed in [8]. The constraint is chosen as a penalty on the deviation from a set of target impulse responses in the bright zone $\mathbf{p}_T$. This leads to the cost function given by [9]

$$J_1(\mathbf{w}) = \beta\mathbf{w}^T\mathbf{R}_D\mathbf{w} + (1-\beta)[(\mathbf{H}_B\mathbf{w}-\mathbf{p}_T)^T(\mathbf{H}_B\mathbf{w}-\mathbf{p}_T)] + \delta\mathbf{w}^T\mathbf{R}\mathbf{w}, \tag{10}$$

where the scalar $\beta \in [0, 1[$ adjusts the trade-off between achieving the desired impulse response in the bright zone and cancelling the sound in the dark zone. The last term is a penalty on the weighted sum of the squared filter coefficients. The scalar $\delta > 0$ adjusts the penalty on the filtered and squared entries in the filter vector $\mathbf{w}$. To increase the penalty for frequencies outside the operating frequency range of the loudspeakers, a weighting matrix $\mathbf{R}$ is included. This block diagonal matrix is written as

$$\mathbf{R} = \begin{bmatrix} \mathbf{B}^T\mathbf{B} & 0 & \cdots & 0 \\ 0 & \mathbf{B}^T\mathbf{B} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{B}^T\mathbf{B} \end{bmatrix}, \tag{11}$$

where $\mathbf{B}$ is a convolution matrix with the frequency weighting implemented as a FIR filter $\mathbf{b} = [b(0), \cdots, b(J-1)]^T$ with $J \leq I$, which defines $\mathbf{B}$ as

$$\mathbf{B} = \begin{bmatrix} b(0) & 0 & 0 \\ \vdots & \ddots & 0 \\ b(J-1) & \ddots & b(0) \\ 0 & \ddots & \vdots \\ 0 & 0 & b(J-1) \end{bmatrix}. \tag{12}$$

In the current paper, it is of interest to extend this cost-function to include an additional penalty in order to control the shape of the resulting FIR filters. This penalty is introduced to enforce a specific filter shape. The envelope penalty can be introduced as the inner product of $\mathbf{w}$ with entries weighted according to the desired filter envelope. The envelope weighting for each filter is $\mathbf{r}_l = [r_l(0), \cdots, r_l(I-1)]^T$, with the concatenated weighting vector $\mathbf{r} = [\mathbf{r}_1^T, \cdots, \mathbf{r}_L^T]^T$ and the total penalty weighting matrix is

$$\mathbf{R}_w = \mathrm{diag}(\mathbf{r}). \tag{13}$$

The proposed cost function is thus

$$J_2(\mathbf{w}) = J_1(\mathbf{w}) + \delta_w\mathbf{w}^T\mathbf{R}_w\mathbf{w}, \tag{14}$$

and the penalty of the envelope shape is controlled by the weighting vector $\mathbf{r}$ and the scalar $\delta_w > 0$, which adjusts the emphasis in the cost-function on the filter shape, relative to the other terms. According to [9] the cost-function in eq. 14 has a global minimum at the concatenated filter vector

$$\mathbf{w}_{2,\min} = \underset{\mathbf{w}}{\mathrm{argmin}}\ J_2(\mathbf{w}) = [\beta\mathbf{R}_D + (1-\beta)\mathbf{R}_B + \delta\mathbf{R} + \delta_w\mathbf{R}_w]^{-1}(1-\beta)\mathbf{H}_B^T\mathbf{p}_T, \tag{15}$$

if $[\beta\mathbf{R}_D + (1-\beta)\mathbf{R}_B + \delta\mathbf{R} + \delta_w\mathbf{R}_w]$ is positive definite.

3. Results

The investigation presented here is based on a series of measured impulse responses from eight 10" woofers placed in a circle in an acoustically dampened room. The impulse responses from each woofer were measured at 120 microphone positions defining each zone as a 6 by 10 element planar array (with 5 cm interelement distance) at two heights (1.30 m and 1.53 m) above the floor. Measurements were performed using exponential sweeps [10] from 0.1 Hz to 24 kHz with a duration of 2 s at a sampling rate of 48 kHz.

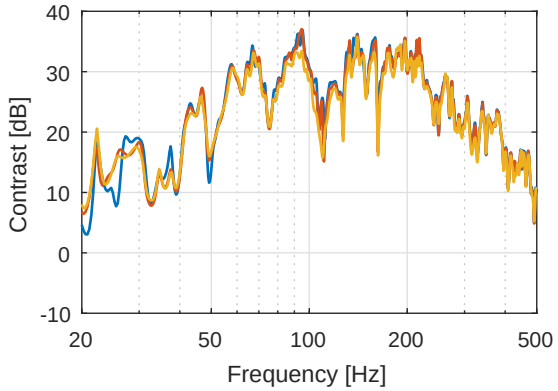
To avoid using the same impulse responses for calculating the filters and evaluating the performance (sometimes referred to as the inverse crime) a number of precautions were taken. The microphone positions in each zone were divided into two different sets: One for calculating the filters and one for evaluating the resulting performance. Likewise, the impulse response measurement was repeated, so one set could be used for determining the filters and a different set could be used for the evaluation according to the evaluation procedure described in [7].

The influence of the filter envelope penalty depends on the desired envelope, as well as the penalty weight. The cost function for controlling the sound field 14 includes a term describing the deviation from a target sound field. To account for the processing and propagation delay in the controlled solution this target sound field usually includes a modelling delay [11]. To avoid filter pre- and post-ringing artifacts it is desired to have short filter responses which rapidly reaches zero, ideally approaching the unit sample sequence. Hence, the desired envelope is unity around the modelling delay and rapidly decreasing towards zero as the time distance to the modelling delay increases. This target can be introduced as a penalty increasing exponentially with the distance from the unit weighting surrounding the modelling delay. Note that the desired envelope might be asymmetric to comply with the different time constants of pre- and post-masking in the auditory system[12]. The envelope penalty weighting of concern can then be expressed as the piecewise function

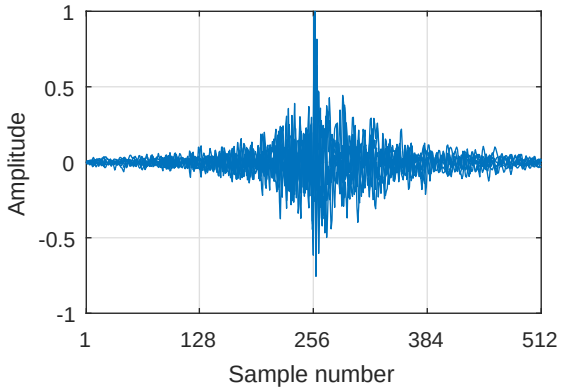
$$r(m) = \begin{cases} \exp\left(\ln(\zeta_l)\frac{m_l - m}{m_l}\right) & m \in [0, m_l] \\ 1 & m \in]m_l, m_u[\\ \exp\left(\ln(\zeta_u)\frac{m - m_u}{M-1-m_u}\right) & m \in [m_u, M-1], \end{cases} \tag{16}$$

where $\delta_w\zeta_l$ and $\delta_w\zeta_u$ are the maximal penalty introduced on pre- and post-ringing, respectively.

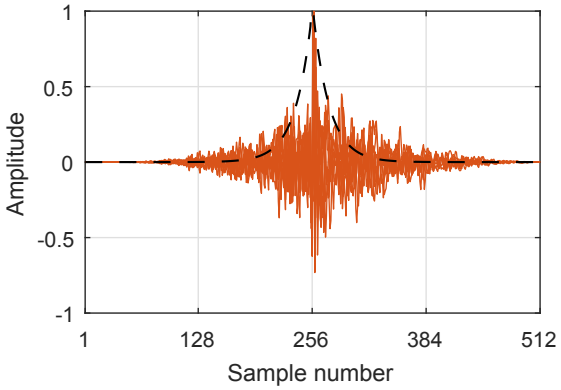
Below, a selected set of results is shown to highlight the potential of the envelope shaping of the FIR filters for the considered scenario. The contrast performance achieved with three different envelope penalties is shown in Fig. 1a. It is seen that the contrast is very similar for the different penalties. However, the filters associated with the contrast plots, shown in Fig. 1b-1d, have different time-domain characteristics. Fig. 1b shows the regular filters determined from minimization of $J_1(\mathbf{w})$ with $\delta = 10^{-4}$. In Fig. 1c the filters are a result of minimizing $J_2(\mathbf{w})$ with a symmetric penalty and unity weighting at 1 sample on each side of the modelling delay, as indicated by the dashed line representing the desired envelope with $\zeta_l = \zeta_u = 10^6$ and $\delta_w = 0.01\delta$. An asymmetric penalty was introduced for the filter shown in Fig. 1d with a unity weighting of 25 samples on each side of the modelling delay with $\zeta_l = 10^{15}$, $\zeta_u = 10^4$, and $\delta_w = \delta$. For all filter calculations, the selected target sound field was the unfiltered response from one of the eight woofers. The results show that without the penalty envelope the filters do not fully reach zero towards the ends. With the symmetric penalty, the filters are forced to zero and with the asymmetric penalty the pre-ringing of the filters is reduced. In Fig. 1c, 1d the reciprocal envelope penalties are plotted on top of the filters to illustrate the desired filter envelopes. It is seen that the symmetric envelope does not entirely match the desired shape, while the asymmetric envelope matches the desired shape well.



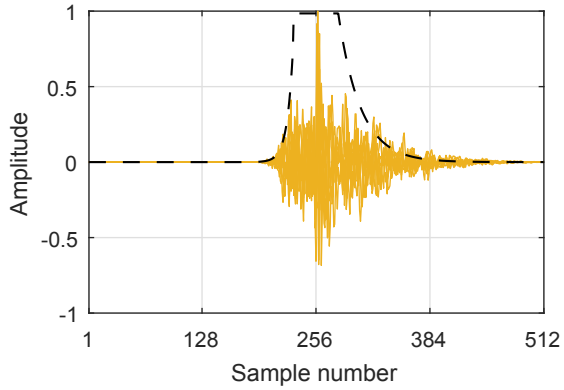
(a) Contrast performance for the filters plotted in Fig. 1b (——), 1c (——), and 1d (——).



(b) (——): Eight superimposed filters determined from minimizing $J_1(\mathbf{w})$.



(c) (——): Eight superimposed filters determined from minimizing $J_2(\mathbf{w})$ with symmetric envelope penalty. (- - -): Reciprocal penalty weight.



(d) (——): Eight superimposed filters determined from minimizing $J_2(\mathbf{w})$ with asymmetric envelope penalty. (- - -): Reciprocal penalty weight.

Figure 1: The resulting contrast performance and the three sets of filters generating the corresponding contrast curves.

4. Discussion

From the results presented in the previous section, it is seen that the envelope shaping of the filter time responses does not compromise the achievable contrast significantly. The filter ringing effects are related to the optimization problem, and might compromise the resulting sound quality perceived by the listener in the bright zone. The optimization problem includes 120 control microphone positions and 8 controllable woofers and is referred to as an over-determined system. As there might not exist a unique solution to the problem, additional constraints or penalties are introduced in the cost-function to obtain an approximate solution with desired envelope properties.

The performance of the introduced envelope penalty can be seen from the comparison of the reciprocal penalty term plotted against the filters in Fig. 1c, 1d. It is shown that the asymmetric filters are well confined by the reciprocal envelope penalty, while the symmetric filters are not. This indicates that the chosen asymmetric penalty is less restrictive than the symmetric, since the latter is not fully realized. It is possible to emphasize the symmetric penalty by increasing δ_w, but that would potentially compromise the acoustic contrast performance. The envelopes shown in this paper are selected to reduce pre- and post-ringing as much as possible without deteriorating the contrast performance. It is possible to choose shorter filter targets, but that is at the cost of reduced contrast between the zones, as short filters cannot represent control of a system with a long impulse response, as is common in listening rooms at low frequencies. In other words, it is possible to adjust the envelope penalty to control the tradeoff between the contrast performance and the resulting envelope

of the filters.

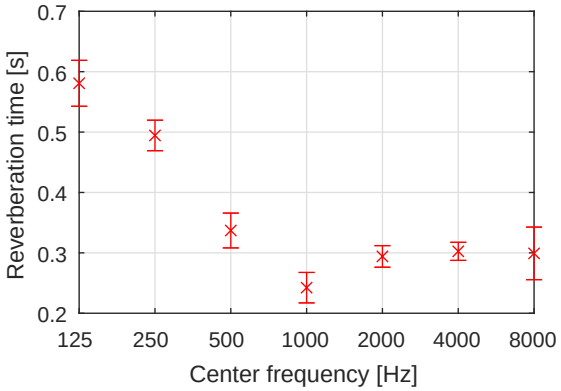
Until now, the underlying assumption has been that the envelope of the filters is directly correlated with the time response of the resulting sound field in the target zone. However, this assumption might not be valid. The simplification is for example violated when the loudspeakers are positioned at different distances to the bright zone. In that case, the desired envelope would penalize loudspeakers with different propagation delays from the majority of the loudspeakers. This facilitates a solution where the filter envelopes are time aligned, even though the loudspeakers have different propagation delays to the bright zone. A solution to this is to choose a specific envelope penalty for each source filter, matching the physical delay from that loudspeaker to the bright zone.

Another question of interest is whether the change in pre- and post-ringing, introduced with the envelope penalty, is perceptually significant. A perceptual study has not been carried out, but the authors seek to provide a brief assessment of whether the change might be perceptually significant, or whether a differently shaped target is required. Multiple factors can be considered, such as: the properties of the reproduction environment and the properties of the human auditory system.

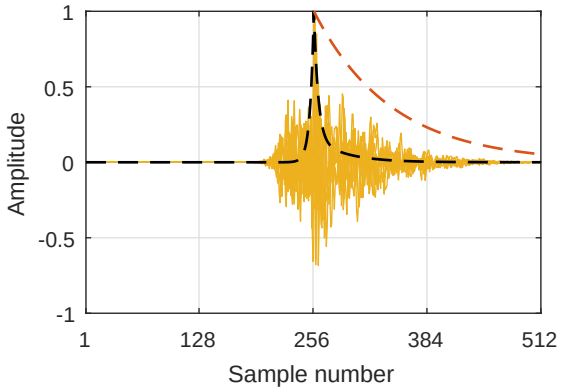
The reverberation time of the room is related to how low the post-ringing of the resulting sound can be. If the reverberation time is long it is not expected that filters with very low post-ringing will be able to control the sound field. The red dashed limit introduced in Fig. 2b is the decay time of the room (0.5 s in the 250 Hz 1 octave band). The reverberation time in the room shown in Fig. 2a is estimated as T_{30} through backwards integration, across three loudspeaker positions and four microphone positions for each loudspeaker positions. By comparing the envelope of the filters with the decay time of the room, it is seen that the envelopes decay faster than the general reverberation. This is related to the woofers being close to the zones, which gives a strong contribution from the direct sound. Therefore, it is more sensible to compare the ringing of the resulting sound field in the bright zone against the response of a woofer without filtering. Convolving the room impulse responses with the asymmetrically penalized filters yields the time-domain responses plotted in Fig. 2c. The decay rate of these resulting responses can be compared with the decay rate of the pressure responses when no envelope penalty is introduced, in Fig. 2d, as well as target sound field in the bright zone shown in Fig. 2e. From this comparison, it can be seen that the filters with and without envelope penalty add both pre- and post-ringing to the resulting pressure responses, relative to the target responses. The main visible difference between the resulting pressure responses, is that the pre-ringing introduced with the envelope penalty is shorter than without it, which is expected from comparison of Fig.1b and 1d. The change in post-ringing is harder to spot from the resulting pressure responses as it is convolved with the room impulse responses. Thus, there is a physical observable change in the responses with and without the envelope penalty, but it is unknown whether this change is perceptually significant.

Whether the pre- and post-ringing of the filters will be audible or not is related to the concept of masking in the auditory system. Masking of complex tones (music) is a separate research field and a thorough analysis and modeling of the audibility of the change is outside the scope of the current work. Instead, the resulting impulse responses are compared with the temporal integration window. In [12] nonsimultaneous masking was explained through loudness integration with a temporal integration window. A window fitted to a single test subject in [12] is plotted on top of the filters with asymmetric envelope penalty in Fig. 2b, and the resulting pressure responses in the bright zone in Fig. 2c and 2d. Thereby, the time constants of this temporal integration window can be compared to the pre- and post-ringing of the asymmetric filters and the pressure responses in the bright zone from filters with and without envelope penalty[1]. From these comparisons, it can be seen that both the filter responses and the pressure responses are longer than this integration window. It is likely that the ringing is perceivable relative to the unit sample sequence (i.e. no introduced ringing influence of either control filters or room reverberation). However, it is not possible to state whether or not the sound field in

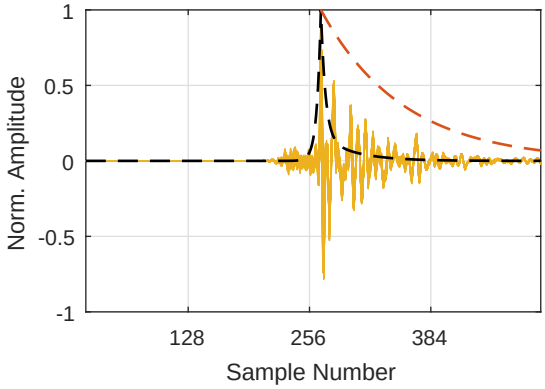
[1]Note that the temporal window has been time-reversed to compare the slope related to post-masking with the post-ringing of the filters.



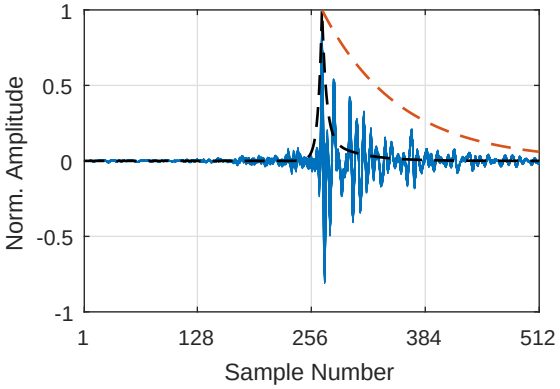
(a) Reverberation time (T_{30}) in octave bands, measured in the listening room.



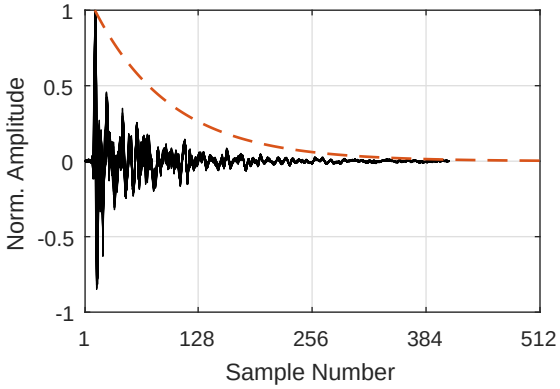
(b) (——): The superimposed asymmetric filters plotted. (- - -): Time-reversed temporal integration window from [12]. (- - -): Exponential decay corresponding to the reverberation time (0.5 s).



(c) (——): The superimposed resulting impulse responses in the bright zone evaluation points, after convolving all loudspeaker responses with the asymmetric penalized filters (Fig. 1d). (- - -): Time-reversed temporal integration window from [12]. (- - -): Exponential decay corresponding to the reverberation time (0.5 s).



(d) (——): The superimposed resulting impulse responses in the bright zone, after convolving all loudspeaker responses with the filters without envelope penalty (Fig. 1b). (- - -): Time-reversed temporal integration window from [12]. (- - -): Exponential decay corresponding to the reverberation time (0.5 s).



(e) (——): The superimposed target (unfiltered response of one of the woofers) impulse responses for all evaluation points in the bright zone. (- - -): Exponential decay corresponding to the reverberation time (0.5 s).

the bright zone, generated by the proposed asymmetric filters, are perceptually different relative to filters with no envelope penalty. For this purpose, a dedicated perceptual experiment is necessary. Such experiment would allow investigations of perceptual thresholds for pre- and post-ringing, which

could be used to determine a trade-off between sound quality and contrast performance.

5. Conclusion

In this paper, it has been shown that the envelope of sound zoning filters can be controlled by introducing an additional penalty term to existing time-domain cost functions. It was seen that it is possible to reduce the pre- and post-ringing of the filters without compromising the resulting acoustic separation between the bright and dark zone.

6. Further Work

From the presented results, it was not possible to argue for a perceptually significant change of the sound quality. Perceptual experiments could be designed to establish thresholds for audible changes in the sound quality resulting from reduced pre- and post-ringing of the filters.

REFERENCES

1. Druyvesteyn, W. F. and Garas, J. Personal Sound, *J. Audio Eng. Soc.*, **45** (9), 685–701, (1997).
2. Choi, J. and Kim, Y. Generation of an Acoustically Bright Zone with an Illuminated Region using Multiple Sources, *J. Acoust. Soc. Am.*, **111** (4), 1695–1700, (2002).
3. Francombe, J., Mason, R., Dewhirst, M., and Bech, S. Elicitation of Attributes for the Evaluation of Audio-on-Audio Interference, *J. Acoust. Soc. Am.*, **136** (5), 2630–2641, (2014).
4. Rämo, J., Marsh, S., Bech, S., Mason, R., and Jensen, S. H., Validation of a Perceptual Distraction Model in a Complex Personal Sound Zone System, *Audio Engineering Society Convention 141*, Los Angeles, USA, 29 September – 2 October, (2016).
5. Wu, Y. J. and Abhayapala, T. D. Spatial Multizone Soundfield Reproduction: Theory and Design, in *IEEE Transactions on Audio, Speech, and Language Processing*, **19** (6), 1711–1720, (2011).
6. Donley, J., Ritz, C. and Kleijn, W. B. Improving Speech Privacy in Personal Sound Zones, *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 20–25 March, (2016).
7. Møller, M. B. and Olsen, M. Sound Zones: On Performance Prediction of Contrast Control Methods, *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Contol*, Guildford, United Kingdom, 18–20 July, (2016).
8. Schellekens, D. H. M., Møller, M. B. and Olsen, M. Time Domain Acoustic Contrast Control Implementation of Sound Zones for Low-Frequency Input Signals, *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 20–25 March, (2016).
9. Gàlvez, M. F. S., Elliott, S. J. and Cheer, J. Time Domain Optimization of Filters Used in a Loudspeaker Array for Personal Audio, in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **23** (11), 1869–1878, (2015).
10. Farina, A. Advancements in Impulse Response Measurements by Sine Sweeps, *Audio Engineering Society Convention 122*, Vienna, Austria, 5–8 May, (2007).
11. Nelson, P. A., Orduña-Bustamante, F. and Hamada, H. Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction, in *IEEE Transactions on Speech and Audio Processing*, **3** (3), 185–192, (1995).
12. Oxenham, A. J. and Moore, B. C. J. Modeling the Additivity of Nonsimultaneous Masking, *Hearing Research*, 80, 105–118, (1994).