DOI: 10.61782/fa.2023.0829

forum acusticum 2023

ROBUST SOUND ZONE FILTERS FOR SYNCHRONIZATION ERRORS

Mo Zhou[1*] **Martin Bo Møller**[2] **Christian Sejer Pedersen**[1]
Niels Evert Marius de Koeijer[2] **Jan Østergaard**[1]
[1] Department of Electronic Systems, Aalborg University, Denmark
[2] Bang & Olufsen, Struer, Denmark

ABSTRACT

Low-frequency sound zones can be created by controlling the sound pressure using loudspeakers distributed throughout the room. When transmitting filtered signals using wireless communication technology, the performance of the low-frequency sound zone system is sensitive to potential synchronization errors among distributed loudspeakers. Recent experiments have shown that a delay of 0.83 ms in only one loudspeaker will lead to a 3-15 dB decrease in the mean acoustic contrast, depending on which loudspeaker is influenced. In this paper, we propose a set of robust filters for the sound zone system by incorporating information about the expected synchronization errors in playback time into the design to increase the robustness towards such errors. With such robust filters, the sound zone system can still achieve comparative performance under more relaxed synchronization requirements among playback from the wireless woofers. The performance of the proposed robust filters is evaluated in a simulated sound zone system with loudspeakers surrounding two control regions. The mean contrasts under various situations for the robust filters are shown to outperform the original filters.

Keywords: *sound zone; synchronization errors; wireless transmission; robustness; contrast.*

**Corresponding author*: *moz@es.aau.dk.*
Copyright: *©2023 Mo Zhou et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

1. INTRODUCTION

Sound zones techniques can provide a scenario where multiple users in the same room can listen to individual playback signals without headphones [1]. This is usually achieved by using an array of loudspeakers to reproduce a signal in the bright zone, defined as the control region where the playback signal is desired, while suppressing the signal in the dark zone. With the principle of superposition, it is then possible to create multiple sound zones [2–4]. Different control schemes are considered for different frequency ranges due to various wavelengths across the audible frequencies [5]. In this paper, we focus on low-frequency sound zone systems, where multiple loudspeakers distributed around the room are often employed.

We consider a sound zone system, where the input signal is first convolved with a set of pre-determined control filters in a server, then transmitted to each loudspeaker to create sound zones. The transmission can be conducted either through a cable network or a wireless network. The portability, flexibility, and lower installation complexity of wireless network motivated us to consider a wireless sound zone system in this paper. However, the possible introduced network errors such as packet losses will undermine the performance of sound zones. A recent experiment in [6] investigated different types of degradation that transmission errors can have on the sound zone performance such as packet loss and showed that mitigation strategies are necessary when network errors occur. Robust filters for low-frequency sound zone systems are proposed in [7] by incorporating expected packet loss information into the filter design and can improve the contrast when packet losses occur. While packet loss can undermine sound zone performances, synchronization errors between the loudspeakers may lead to a large decrease

in the mean acoustic contrast. It has been shown in [6] that as little as 0.83 ms delay in one loudspeaker can reduce the mean contrast by 3-15 dB, depending on which loudspeaker is influenced. During the last decade, several wireless synchronization approaches have been introduced [8]. However, these attempts could not improve synchronization errors beyond the millisecond range on commodity hardware, and it is even harder in a wireless scenario [9–11]. Therefore, it is crucial to improve the robustness to the synchronization errors in sound zone systems.

In this paper, we consider a low-frequency sound zone systems where each loudspeaker is synchronized with a given reference, i.e. master clock. The audio packet each loudspeaker receives over the network has a timestamp, which indicates that a particular packet should be played when the local clock reaches a certain value. As such, if the local clock of the loudspeaker is not perfectly synchronized with the master clock, the audio will be reproduced slightly before or after it should have been reproduced. To mitigate the effect of such synchronization errors, we develop a set of robust filters given the information about the synchronization accuracy between loudspeakers. We model the clock difference between each loudspeaker's local clock and the master clock by a simple delay, assuming that packet size is small relative to the rate at which the local clock drifts from the master clock. A simulation study is conducted to show that the proposed filters can improve the contrast, and also suggest that the proposed filters can relax the required accuracy of the synchronization to ensure the contrast degradation is negligible.

2. PRELIMINARY

For microphone m $(m = 1, ..., M)$ and loudspeaker l $(l = 1, ..., L)$, we assume that the room impulse responses (RIRs) in the time domain can be represented by $\tilde{\boldsymbol{h}}_{m,l} = (\tilde{h}_{m,l}(0), ..., \tilde{h}_{m,l}(J-1))^T$ and the filters can be written as $\tilde{\boldsymbol{w}}_l = (\tilde{w}_l(0), ..., \tilde{w}_l(I-1))^T$. The sound pressure at time n recorded by microphone m due to loudspeaker l without packet loss can be written as

$$\tilde{p}_{m,l}(n) = \sum_{j=0}^{J-1} \tilde{h}_{m,l}(j) \sum_{i=0}^{I-1} \tilde{w}_l(i) x_s(n-i-j), \quad (1)$$

where x_s is the input audio signal. Assuming the source signal to be spectrally flat, it can be simplified as a unit sample sequence and evaluate the expectation of the squared pressure measured at the observation point. With this assumption, equation (1) can be written as

$$\tilde{\boldsymbol{p}}_{m,l} = \tilde{\boldsymbol{H}}_{m,l} \tilde{\boldsymbol{w}}_l, \quad (2)$$

$$\tilde{\boldsymbol{H}}_{m,l} = \begin{pmatrix} \tilde{h}_{m,l}(0) & & \\ \vdots & \ddots & \\ \tilde{h}_{m,l}(J-1) & \ddots & \tilde{h}_{m,l}(0) \\ & \ddots & \vdots \\ & & \tilde{h}_{m,l}(J-1) \end{pmatrix},$$

where $\tilde{\boldsymbol{H}}_{m,l} \in \mathbb{R}^{N \times I}$ is a convolutional matrix, $\tilde{\boldsymbol{p}}_{m,l} \in \mathbb{R}^N$ and $N = I + J - 1$. We can equivalently write this using a circulant matrix and a matrix $\boldsymbol{Z}$ introducing zero-padding of the filter

$$\tilde{\boldsymbol{p}}_{m,l} = \hat{\boldsymbol{H}}_{m,l} \boldsymbol{Z} \tilde{\boldsymbol{w}}_l, \quad (3)$$

$$\hat{\boldsymbol{H}}_{m,l} = \begin{pmatrix} \tilde{h}_{m,l}(0) & \dots & \tilde{h}_{m,l}(1) \\ \vdots & \ddots & \vdots \\ \vdots & & \tilde{h}_{m,l}(J-1) \\ \tilde{h}_{m,l}(J-1) & \ddots & \vdots \\ 0 & \ddots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \tilde{h}_{m,l}(0) \end{pmatrix},$$

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{I}_{I \times I} \\ \boldsymbol{0}_{(J-1) \times I} \end{pmatrix},$$

where $\hat{\boldsymbol{H}}_{m,l} \in \mathbb{R}^{(I+J-1) \times (I+J-1)}$, $\boldsymbol{I}_{I \times I}$ is an $I \times I$ identity matrix and $\boldsymbol{0}_{(J-1) \times I}$ is a $(J-1) \times I$ matrix with all 0's. The control filters $\tilde{\boldsymbol{w}}_l$ in equation (3) are often derived by solving an optimization problem so that the reproduced signals $\tilde{\boldsymbol{p}}_{m,l}$ in each zone are close to the desired playback sound pressure [1].

3. MODEL WITH SYNCHRONIZATION ERRORS

Assume that we have circular delay τ_l for loudspeaker l measured in samples. Since we can diagonalize the circulant matrix using the discrete Fourier transform matrix $\boldsymbol{F}$ with $F_{n,n'} = e^{-(n-1)(n'-1)\frac{2\pi i}{N}}, n, n' = 1, ..., N$, we can

express the pressure in the frequency domain as

$$\begin{aligned}
&\boldsymbol{p}_{m,l} = \boldsymbol{\Gamma}_l \boldsymbol{H}_{m,l} \boldsymbol{F} \boldsymbol{Z} \tilde{\boldsymbol{w}}_l, \qquad (4)\\
&\boldsymbol{p}_{m,l} = (p_{m,l}(f_1), ..., p_{m,l}(f_N))^T \text{ with } f_n = n/N,\\
&\boldsymbol{\Gamma}_l = \text{diag}(\boldsymbol{\tau}_l), \boldsymbol{\tau}_l = (e^{-2\pi i f_1 \tau_l}, ..., e^{-2\pi i f_N \tau_l})^T,\\
&\boldsymbol{H}_{m,l} = \text{diag}(\boldsymbol{F} \begin{pmatrix} \tilde{\boldsymbol{h}}_{m,l} \\ \boldsymbol{0}_{I-1} \end{pmatrix}) = \text{diag}(\boldsymbol{h}_{m,l}),\\
&\text{with } \boldsymbol{h}_{m,l} = (h_{m,l}(f_1), ..., h_{m,l}(f_N))^T.
\end{aligned}$$

Denote

$$\begin{aligned}
&\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_{1,1} & \dots & \boldsymbol{H}_{1,L} \\ \vdots & \ddots & \vdots \\ \boldsymbol{H}_{M,1} & \dots & \boldsymbol{H}_{M,L} \end{pmatrix} \in \mathbb{C}^{NM \times NL},\\
&\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\tau}_1^T, ..., \boldsymbol{\tau}_L^T)^T \in \mathbb{C}^{NL \times NL},\\
&\tilde{\boldsymbol{w}} = (\tilde{\boldsymbol{w}}_1^T, ..., \tilde{\boldsymbol{w}}_L^T)^T \in \mathbb{C}^{IL},
\end{aligned}$$

the sound pressure for M microphone positions can be written as

$$\boldsymbol{p} = \boldsymbol{H}\boldsymbol{\Gamma}(\boldsymbol{I}_L \otimes \boldsymbol{F}\boldsymbol{Z})\tilde{\boldsymbol{w}} = \boldsymbol{H}\boldsymbol{\Gamma}\boldsymbol{D}\tilde{\boldsymbol{w}}, \qquad (5)$$

where $\boldsymbol{p} = (\boldsymbol{p}_1^T, ..., \boldsymbol{p}_M^T)^T \in \mathbb{C}^{NM}$ with $\boldsymbol{p}_m = \sum_{l=1}^{L} \boldsymbol{H}_{m,l}\boldsymbol{\Gamma}_l \boldsymbol{F}\boldsymbol{Z}\tilde{\boldsymbol{w}}_l$, $\otimes$ represents the Kronecker product and $\boldsymbol{D} = \boldsymbol{I}_L \otimes \boldsymbol{F}\boldsymbol{Z}$. Therefore, the sound pressure for M microphone positions in the bright and dark zones can be written as

$$\boldsymbol{p}_B = \boldsymbol{H}_B\boldsymbol{\Gamma}\boldsymbol{D}\tilde{\boldsymbol{w}}, \quad \boldsymbol{p}_D = \boldsymbol{H}_D\boldsymbol{\Gamma}\boldsymbol{D}\tilde{\boldsymbol{w}}.$$

4. FILTER DERIVATION WITH SYNCHRONIZATION ERRORS

As we observe in the considered sound zone system, the delay values between loudspeakers are mostly concentrated within a small range, with some outliers from time to time. Their statistical distribution usually exhibits heavy-tailed properties, i.e. asymmetric and right-skewed in the sense that the long tail is on the right side of the distribution and extremely large values occur frequently. Applications of heavy-tailed distribution in the delay of network transmission have also been researched in [12–14]. Levy distribution is commonly used to characterize heavy tail property. Assume that the time delay τ_l follows from unshifted Levy distribution with scale parameter c_l:

$$\tau_l \sim \text{Levy}(c_l), \text{with } f(\tau_l; c_l) = \sqrt{\frac{c_l}{2\pi}} \frac{e^{-\frac{c_l}{2\tau_l}}}{\tau_l^{3/2}}, \tau_l > 0.$$

In Figure 1 (a), we give some probability density functions of Levy distribution with different scale parameters c_l, note that one sample delay equals to 0.083ms in our low-frequency sound zone system with sampling frequency 1200 Hz. The scale parameter c_l characterizes the dispersion of delay values, i.e. if c_l is small, the delay values will be more concentrated. Figure 1 (b) also plots the corresponding cumulative distribution functions as well as a dotted line indicating the median. We only define the delay on the domain $\tau_l > 0$, excluding the case of perfect synchronization. In Figure 1 (c), we also show a histogram of measured delays in the wireless transmission between two Linux laptops in a preliminary test as well as a fitted Levy distribution, proving that Levy distribution is suitable for describing delay distribution in our system.

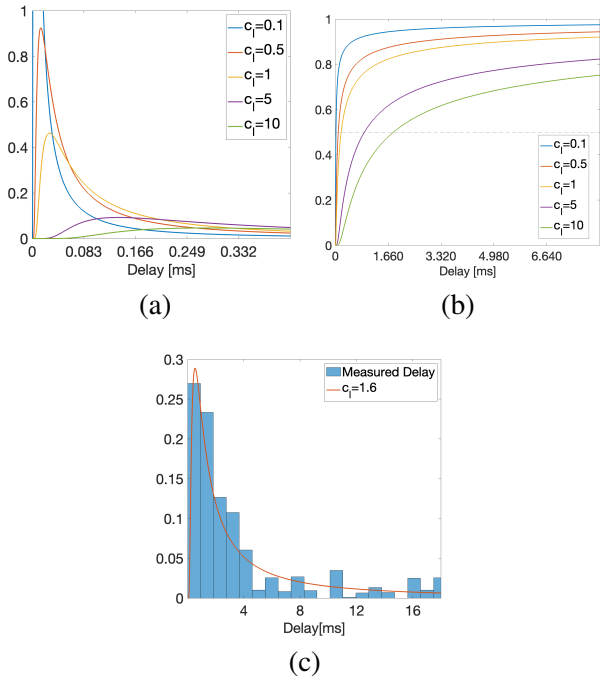


Figure 1: (a) Probability density functions of Levy distribution. (b) Cumulative distribution functions of Levy distribution. (c) Histogram of measured delay and fitted Levy distribution with $c_l = 1.6$.

Let $\boldsymbol{p}_T = (\boldsymbol{p}_{T,1}^T, ..., \boldsymbol{p}_{T,M}^T)^T \in \mathbb{C}^{NM}$ be the desired sound pressure at frequencies $f_1, ..., f_N$ for M microphone positions, where $\boldsymbol{p}_{T,m} = (p_{T,m}(f_1), ..., p_{T,1}(f_N))^T$. The cost function that minimizes the reproduction error for a given set

of filters can be written as

$$J(\boldsymbol{w}) = (1-\beta)\mathbb{E}\|\boldsymbol{p}_T - \boldsymbol{H}_B\boldsymbol{\Gamma}\boldsymbol{D}\tilde{\boldsymbol{w}}\|^2 + \beta\mathbb{E}\|\boldsymbol{H}_D\boldsymbol{\Gamma}\boldsymbol{D}\tilde{\boldsymbol{w}}\|^2 + \lambda_w\tilde{\boldsymbol{w}}^H\boldsymbol{R}_w\tilde{\boldsymbol{w}} + \delta\tilde{\boldsymbol{w}}^H\tilde{\boldsymbol{w}}, \quad (6)$$

where $\beta \in [0,1]$ adjusts the trade-off between achieving the desired impulse response in the bright zone and supressing the sound in the dark zone. The expectation $\mathbb{E}$ is with respect to the delay $\boldsymbol{\Gamma}$, $\boldsymbol{R}_w$ is a weighting matrix for controlling the shape of the resulting filters as suggest in [15], δ is to limit the gain of the loudspeakers. The robust filters $\boldsymbol{w}_{opt}$ can be estimated by minimizing (6) and has the following form:

$$\boldsymbol{w}_{opt} = [(1-\beta)\boldsymbol{D}^H(\boldsymbol{H}_B^H\boldsymbol{H}_B \odot \Omega)\boldsymbol{D} + \beta\boldsymbol{D}^H(\boldsymbol{H}_D^H\boldsymbol{H}_D \odot \Omega)\boldsymbol{D} + \lambda_w\boldsymbol{R}_w + \delta\boldsymbol{I}_{IL\times IL}]^{-1}(1-\beta)\boldsymbol{D}^H\boldsymbol{\Psi}\boldsymbol{H}_B^H\boldsymbol{p}_T, \quad (7)$$

where $\odot$ denotes the Hadamard product and $\Omega, \boldsymbol{\Psi}$ are defined in the appendix, where the details of the derivation of $\boldsymbol{w}_{opt}$ are also given.

5. SIMULATION

5.1 Simulation Setting

We simulate a 5.5 m by 8.65 m by 2.7 m room using Green's function for point sources in rectangular rooms, with 0.6s T_{60} reverberation time and $L = 8$ loudspeakers. The number of microphone positions sampled in the bright and dark zones are chosen as $M_B = M_D = 75$. The setup is illustrated in Figure 2. The RIRs and the filters are of length $J = 600$ and $I = 300$, respectively. In addition, we take $\beta = 0.97$ and $\lambda_w = 10^{-7}$ in the cost function for all evaluations. The weighting matrix $\boldsymbol{R}_w$ is chosen according to [15]. We focus on the cases where only one loudspeaker is subject to synchronization errors. Denote $\boldsymbol{\omega}_{l,c}, l = 1, ..., 8$ as our proposed filters derived by assuming loudspeaker l has a time delay $\tau_l \sim$ Levy(c), we vary the scale parameters c as $c = 0.1, 0.5, 1, 5, 10$ to see how these filters behave.

5.2 Contrast

We evaluate the acoustic contrast of our proposed filters $\boldsymbol{\omega}_{l,c}, l = 1, ..., 8$ in this section. To isolate the influence of a particular input signal, and to avoid the influence of the upsampling procedure when evaluating fractional delays, we use equation (4) to calculate sound pressures in the frequency domain for each zone and evaluate the performance of $\boldsymbol{\omega}_{l,c}$ directly. More specifically, for each $\boldsymbol{\omega}_{l,c}$,

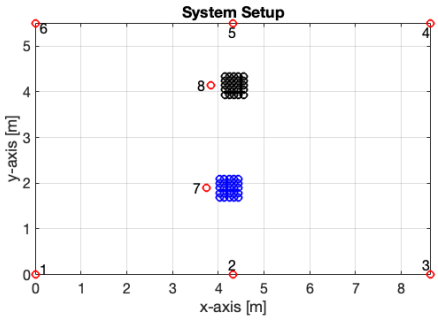


Figure 2: System setup for the simulation. The blue and black circles are the microphones in the bright and dark zones respectively. The red circles are the loudspeakers.

we draw 100 random samples from Levy(c) as the introduced delay in loudspeaker l. The sound pressures for each realization in the bright and dark zones are calculated using equation (4), then the Mean Contrast for each realization is calculated by the formula (8):

$$MC = 10 \times \log_{10}\left(\frac{ME_B}{ME_D}\right), \quad (8)$$

$$ME_B = \sum_{n=1}^{N}\sum_{m=1}^{M}(\sum_{l=1}^{L}\hat{p}_{B,m,l}(f_n))^2/M_BN,$$

$$ME_D = \sum_{n=1}^{N}\sum_{m=1}^{M}(\sum_{l=1}^{L}\hat{p}_{D,m,l}(f_n))^2/M_DN,$$

where $\hat{p}_{B,m,l}, \hat{p}_{D,m,l}$ are the reproduced sound pressures in the bright and dark zones respectively using equation (4) with $\tilde{\boldsymbol{w}}_l$ substituted by the estimated filters. The same realization is also applied to the original filters (denoted by $\boldsymbol{\omega}_{old}$) to calculate the Mean Contrast for comparison. We average across all 100 realizations for each $\boldsymbol{\omega}_{l,c}$ and $\boldsymbol{\omega}_{old}$. Figure 3 gives the average Mean Contrast for $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{l,c}$ under different scale parameters c. Generally, $\boldsymbol{\omega}_{l,c}$ outperforms $\boldsymbol{\omega}_{old}$ and can provide 10 dB higher contrast, depending on which loudspeaker is affected. With scale parameter c increases, i.e. higher probability of larger delay, contrasts of $\boldsymbol{\omega}_{old}$ decrease while $\boldsymbol{\omega}_{l,c}$ can still give robust high contrasts. Table 1 also give the average Mean Contrasts over 100 realizations for $\boldsymbol{\omega}_{l,c}, l = 1, ..., 8, c = 0.1, 0.5, 1, 5, 10$ and $\boldsymbol{\omega}_{old}$, as well as the corresponding standard deviations in the brackets. It can be seen that the standard deviations for both filters decrease when the width of the underlying distribution increases, while our

proposed filters have a much smaller standard deviations and therefore are more robust.

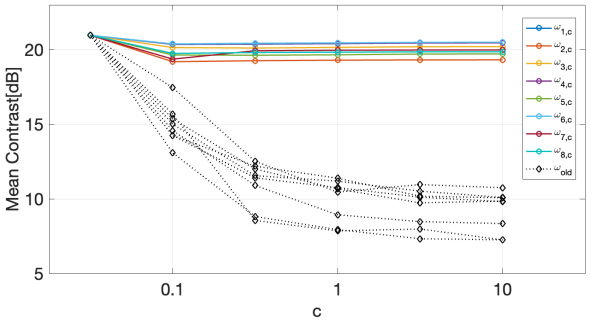


Figure 3: Average Mean Contrast for ω_{old} and $\omega_{l,c}$ under different scale parameters c. Solid lines represent average Mean Contrast for $\omega_{l,c}$ when evaluated with a delay generated from Levy(c) in loudspeaker l . $\cdots\diamond\cdots$ represent avearge Mean Contrast for ω_{old} when evaluated with a delay generated from Levy(c) in loudspeaker l.

To evaluate the behaviors of our proposed filters at different frequencies, Figure 4 plots the frequency contrast of ω_{old} and $\omega_{1,c}$ when evaluated with a delay $\tau_1 = c$ in loudspeaker 1 [1] . Note that when the delay $\tau_1 = 0.1, 1, 10$, it is equivalent to 0.083, 0.83 and 8.3 ms in our case. When the synchronization error is small, e.g. 0.083 ms, contrasts of ω_{old} and $\omega_{1,c}$ are close across all frequency ranges in 20-500 Hz. When the synchronization error increases, $\omega_{1,c}$ is more robust and gives higher contrasts, especially at 20-300 Hz. We also plot contrasts of ω_{old} and $\omega_{1,10}$ when evaluated without delay in Figure 4 (c) [2] , showing that the proposed filters can provide similar contrast as ω_{old} when there is no delay.

[1] Due to the limitation of space, we only present results for loudspeaker 1 and $c = 0.1, 1, 10$ here.

[2] For $c = 0.1, 1$, the contrasts evaluated without delay are very close to the contrasts evaluated with delay, thus we only show the result for $c = 10$.

Table 1: The average Mean Contrasts [dB] over 100 realizations for $\omega_{l,c}, l = 1, ..., 8, c = 0.1, 0.5, 1, 5, 10$ and ω_{old}, with the corresponding standard deviations in the brackets.

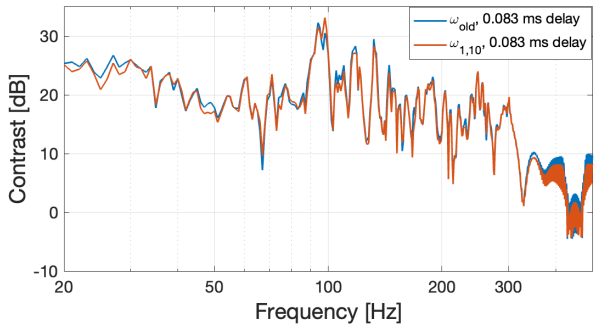
	Loudspeaker 1		Loudspeaker 2	
	ω_{old}	$\omega_{1,c}$	ω_{old}	$\omega_{2,c}$
$c = 0.1$	15.4 (4.31)	20.4 (0.14)	13.1 (4.90)	19.2 (0.32)
$c = 0.5$	12.0 (1.85)	20.4 (0.02)	8.8 (2.76)	19.3 (0.05)
$c = 1$	10.7 (3.16)	20.4 (0.01)	7.9 (1.97)	19.3 (0.03)
$c = 5$	9.7 (1.30)	20.4 (0.00)	7.3 (0.79)	19.3 (0.01)
$c = 10$	9.9 (0.94)	20.5 (0.00)	7.3 (0.77)	19.3 (0.00)
	Loudspeaker 3		Loudspeaker 4	
	ω_{old}	$\omega_{3,c}$	ω_{old}	$\omega_{4,c}$
$c = 0.1$	14.3 (3.94)	20.1 (0.23)	17.5 (3.46)	20.4 (0.09)
$c = 0.5$	11.4 (2.42)	20.1 (0.03)	12.5 (2.24)	20.4 (0.02)
$c = 1$	10.8 (1.83)	20.2 (0.03)	10.5 (1.20)	20.4 (0.01)
$c = 5$	10.1 (1.13)	20.2 (0.01)	11.0 (1.72)	20.4 (0.00)
$c = 10$	9.8 (0.85)	20.2 (0.00)	10.7 (1.59)	20.5 (0.00)
	Loudspeaker 5		Loudspeaker 6	
	ω_{old}	$\omega_{5,c}$	ω_{old}	$\omega_{6,c}$
$c = 0.1$	14.6 (4.05)	19.6 (0.32)	14.3 (4.90)	20.4 (0.12)
$c = 0.5$	10.9 (3.02)	19.6 (0.06)	12.2 (2.63)	20.4 (0.02)
$c = 1$	8.9 (1.49)	19.7 (0.03)	11.4 (2.14)	20.5 (0.01)
$c = 5$	8.5 (0.67)	19.7 (0.01)	10.2 (1.02)	20.5 (0.00)
$c = 10$	8.3 (0.64)	19.7 (0.00)	10.1 (1.00)	20.5 (0.00)
	Loudspeaker 7		Loudspeaker 8	
	ω_{old}	$\omega_{7,c}$	ω_{old}	$\omega_{8,c}$
$c = 0.1$	15.7 (5.30)	19.4 (0.92)	15.0 (3.72)	19.7 (0.36)
$c = 0.5$	8.6 (2.57)	19.9 (0.09)	11.6 (2.70)	19.8 (0.06)
$c = 1$	7.9 (1.53)	20.0 (0.05)	11.2 (1.70)	19.8 (0.03)
$c = 5$	8.0 (2.03)	20.0 (0.02)	10.5 (1.21)	19.9 (0.01)
$c = 10$	7.3 (0.62)	20.0 (0.01)	10.1 (0.73)	19.9 (0.00)

6. DISCUSSION

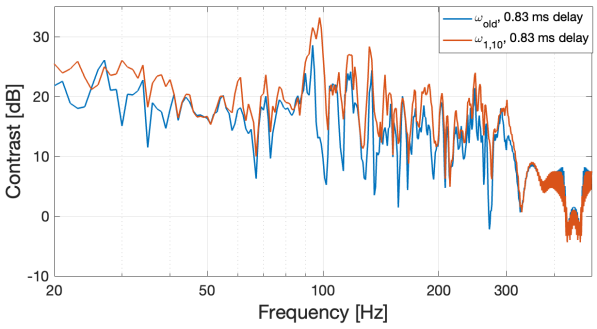
From the contrast result presented in the previous section, incorporating information of expected synchronization errors in playback time into the design of the filters can improve the performance of the wireless low-frequency sound zone systems. The average Mean Contrast performance based on the repeated random realizations in Figure 3 shows that our proposed filters are more robust to the synchronization errors than the original filters, and the improvement is more significant for large delays. The improvement of the contrast at different frequencies can also be seen from an example of $\boldsymbol{\omega}_{1,c}$ in Figure 4.

The mean contrast result also suggests the required accuracy of the synchronization of the sound zone system when the contrast decrement is negligible. When we only use the original filters, no more than 0.05 ms delay is allowed to ensure less than 2 dB decrement of the mean contrast, which is strict in practice. The proposed filters can relax this requirement due to its robustness.

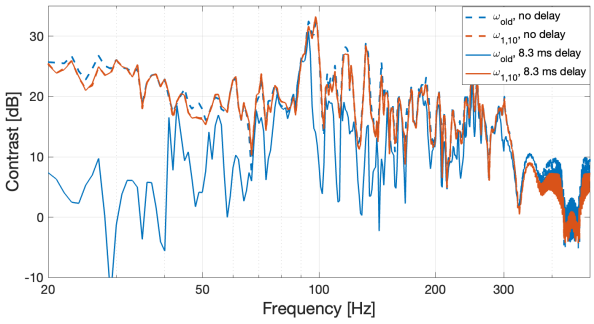
In addition, we have to notice that different loudspeakers have different degrees of sensitivity to the synchronization errors, depending on the locations of these loudspeakers. And the improvement of incorporating the synchronization errors in different loudspeakers varies from loudspeaker to loudspeaker. For example, loudspeaker 7 is most sensitive to the synchronization errors since it is closest to the bright zone, with largest decrement when synchronization errors happen if $\boldsymbol{\omega}_{old}$ are used. $\boldsymbol{\omega}_{7,c}$ can reduce this decrement due to the synchronization errors and gives more robust performance.



(a) $c = 0.1$



(b) $c = 1$



(c) $c = 10$

Figure 4: Contrast of $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{1,c}$ when evaluated with a synchronization error c in loudspeaker 1 ((a)-(c)) and without delay (c).

7. CONCLUSION

This paper proposes a set of robust filters for wireless low-frequency sound zone systems by incorporating information on expected synchronization errors in playback time into the design. The proposed robust filters can improve the contrast when synchronization errors occur compared with the original filters, and also relax the required accuracy. Further investigation will consider dependent synchronization error cases and testing the proposed filters in real-world wireless sound zone systems.

8. APPENDIX

The robust filters $\boldsymbol{w}_{opt}$ can be derived from (6) with

$$\begin{aligned}\boldsymbol{w}_{opt} &= [(1-\beta)\boldsymbol{D}^H\mathbb{E}(\boldsymbol{\Gamma}^H\boldsymbol{H}_B^H\boldsymbol{H}_B\boldsymbol{\Gamma})\boldsymbol{D} \\ &\quad +\beta\boldsymbol{D}^H\mathbb{E}(\boldsymbol{\Gamma}^H\boldsymbol{H}_D^H\boldsymbol{H}_D\boldsymbol{\Gamma})\boldsymbol{D}+\lambda_w\boldsymbol{R}_w \\ &\quad +\delta\boldsymbol{I}_{IL\times IL}]^{-1}(1-\beta)\boldsymbol{D}^H(\mathbb{E}(\boldsymbol{\Gamma}^H))\boldsymbol{H}_B^H\boldsymbol{p}_T,\end{aligned} \tag{9}$$

where $(\cdot)^H$ denotes the Hermitian transpose. In (9), $\boldsymbol{\Gamma}^H = \text{diag}(\boldsymbol{\tau}_1^H,...,\boldsymbol{\tau}_L^H)$ where $\boldsymbol{\tau}_l^H = (e^{2\pi i f_1\tau_l},...,e^{2\pi i f_N\tau_l}), l = 1,...,L$. Since the characteristic function of a random variable $X \sim \text{Levy}(c)$ is $\mathbb{E}(e^{itX}) = e^{-\sqrt{-2ict}}$, we have $\mathbb{E}(e^{2\pi i f_n\tau_l}) = e^{-\sqrt{-4\pi i c_l f_n}}, n = 1,...,N$. Therefore, $\Psi_l = \mathbb{E}(\boldsymbol{\tau}_l^H) = (\mathbb{E}(e^{2\pi i f_1\tau_l}),...,\mathbb{E}(e^{2\pi i f_N\tau_l})) = (e^{-\sqrt{-4\pi i c_l f_1}},...,e^{-\sqrt{-4\pi i c_l f_N}})$, and $\boldsymbol{\Psi} = \mathbb{E}(\boldsymbol{\Gamma}^H) = \text{diag}(\Psi_1,...,\Psi_L)$.

For the expectation of the second moment term, each block of $\mathbb{E}(\boldsymbol{\Gamma}^H\boldsymbol{H}_B^H\boldsymbol{H}_B\boldsymbol{\Gamma})$ has the form

$$\begin{aligned}&[\mathbb{E}(\boldsymbol{\Gamma}^H\boldsymbol{H}_B^H\boldsymbol{H}_B\boldsymbol{\Gamma})]_{l_1,l_2} \\ &= \sum_{m=1}^{M}\mathbb{E}(\boldsymbol{\Gamma}_{l_1}^H\boldsymbol{H}_{B,m,l_1}^H\boldsymbol{H}_{B,m,l_2}\boldsymbol{\Gamma}_{l_2})),\end{aligned} \tag{10}$$

where $l_1, l_2 = 1,...,L$ and each block is a diagonal matrix. For a given m, the expectation can be written as

$$\begin{aligned}&\mathbb{E}(\boldsymbol{\Gamma}_{l_1}^H\boldsymbol{H}_{B,m,l_1}^H\boldsymbol{H}_{B,m,l_2}\boldsymbol{\Gamma}_{l_2}) \\ &= \text{diag}(\overline{h_{B,m,l_1}(f_1)}h_{B,m,l_2}(f_1)\mathbb{E}(e^{2\pi i f_1(\tau_{l_1}-\tau_{l_2})}), \\ &\quad ...,\overline{h_{B,m,l_1}(f_N)}h_{B,m,l_2}(f_N)\mathbb{E}(e^{2\pi i f_N(\tau_{l_1}-\tau_{l_2})}),\end{aligned} \tag{11}$$

where $\overline{(\cdot)}$ denotes the complex conjugation. If $l_1 = l_2 = l$, equation (11) becomes

$$\text{diag}(\overline{h_{B,m,l}(f_1)}h_{B,m,l}(f_1),...,\overline{h_{B,m,l}(f_N)}h_{B,m,l}(f_N))),$$

and (10) becomes $\sum_{m=1}^{M}\boldsymbol{H}_{B,m,l}^H\boldsymbol{H}_{B,m,l}$. If $l_1 \neq l_2$, assume that the delay in different loudspeakers are independent with respect to each other, i.e. τ_{l_1},τ_{l_2} are independent, the entries in equation (11) become

$$\begin{aligned}&\overline{h_{B,m,l_1}(f_n)}h_{B,m,l_2}(f_n)\mathbb{E}(e^{2\pi i f_n\tau_{l_1}})\mathbb{E}(e^{-2\pi i f_n\tau_{l_2}}) \\ &= \overline{h_{B,m,l_1}(f_n)}h_{B,m,l_2}(f_n)e^{-\sqrt{-4\pi i c_{l_1} f_n}}e^{-\sqrt{4\pi i c_{l_2} f_n}},\end{aligned}$$

and (10) is therefore

$$\Omega_{l_1,l_2}\sum_{m=1}^{M}\boldsymbol{H}_{B,m,l_1}^H\boldsymbol{H}_{B,m,l_2},$$

where

$$\begin{aligned}\Omega_{l_1,l_2} &= \text{diag}(e^{-\sqrt{-4\pi i c_{l_1} f_1}}e^{-\sqrt{4\pi i c_{l_2} f_1}}, \\ &\quad ...,e^{-\sqrt{-4\pi i c_{l_1} f_N}}e^{-\sqrt{4\pi i c_{l_2} f_N}}).\end{aligned}$$

Thus, the expectation of the second moment term can be written as

$$\begin{aligned}&\mathbb{E}(\boldsymbol{\Gamma}^H\boldsymbol{H}_B^H\boldsymbol{H}_B\boldsymbol{\Gamma}) \\ &= \boldsymbol{H}_B^H\boldsymbol{H}_B\odot\begin{pmatrix}1 & \cdots & \Omega_{1,L} \\ \Omega_{2,1} & \cdots & \Omega_{2,L} \\ \vdots & \ddots & \vdots \\ \Omega_{L,1} & \cdots & 1\end{pmatrix} \\ &:= \boldsymbol{H}_B^H\boldsymbol{H}_B\odot\Omega.\end{aligned} \tag{12}$$

Similarly, for the dark zone,

$$\mathbb{E}(\boldsymbol{\Gamma}^H\boldsymbol{H}_D^H\boldsymbol{H}_D\boldsymbol{\Gamma}) = \boldsymbol{H}_D^T\boldsymbol{H}_D\odot\Omega.$$

Thus, the equation (9) has the form

$$\begin{aligned}\boldsymbol{w}_{opt} &= [(1-\beta)\boldsymbol{D}^H(\boldsymbol{H}_B^H\boldsymbol{H}_B\odot\Omega)\boldsymbol{D} \\ &\quad +\beta\boldsymbol{D}^H(\boldsymbol{H}_D^H\boldsymbol{H}_D\odot\Omega)\boldsymbol{D}+\lambda_w\boldsymbol{R}_w \\ &\quad +\delta\boldsymbol{I}_{IL\times IL}]^{-1}(1-\beta)\boldsymbol{D}^H\boldsymbol{\Psi}\boldsymbol{H}_B^H\boldsymbol{p}_T.\end{aligned} \tag{13}$$

9. ACKNOWLEDGEMENTS

This work is partly funded by the Innovation Fund Denmark (IFD) under File No. 9069-00038B in the project: Interactive Sound Zones for Better Living (ISOBEL).

10. REFERENCES

[1] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Processing Magazine*, vol.32, no. 2, pp. 81–91, 2015.

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

[3] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen, and J. A. Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *The Journal of the Acoustical Society of America*, vol.135, no. 4, pp. 1929–1940, 2014.

[4] P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen, "Personal audio with planar bright zone," *The Journal of the Acoustical Society of America*, vol.136, no. 4, pp. 1725–1735, 2014.

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

[6] C. S. Pedersen, M. B. Møller, and J. Østergaard "Effect of wireless transmission errors on sound zone performance at low frequencies," *EUROREGIO BNAM2022 Joint Acoustic Conference*, pp. 115–124, 2022.

[7] M. Zhou, M. B. Møller, C. S. Pedersen, and J. Østergaard, "Robust FIR Filters for Wireless Low-frequency Sound Zones," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2023.

[8] A. Mahmood, R. Exel, H. Trsek, and T. Sauter, "Clock synchronization over IEEE 802.11 - A survey of methodologies and protocols," *IEEE Trans. Ind. Informat.*, 2017.

[9] N. Facchi, F. Gringoli, F. Ricciato, and A. Toma, "Emitter localisation from reception timestamps in asynchronous networks," *Comput. Netw.*, vol. 88, pp. 202–217, 2015.

[10] S. Li, M. Hedley, K. Bengston, D. Humphrey, M. Johnson, and W. Ni, "Passive localization of standard WiFi devices," *IEEE Syst. J.*, vol. 13, no. 4, pp. 3929–3932, 2019.

[11] A. M. Romanov, F. Gringoli, and A. Sikora, "A precise synchronization method for future wireless TSN networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3682-3692, 2020.

[12] D. Muñoz-Rodríguez, S. Villarreal Reyes, C. Vargas Rosales, M. Angulo Bernal, D. Torres-Román, and L. Rizo Domínguez, "Heavy tailed network delay: An alpha-stable," in *Computación y Sistemas*, vol. 10, no. 1, pp. 16–27, 2006.

[13] J. Liebeherr, A. Burchard and F. Ciucu, "Delay Bounds in Communication Networks With Heavy-Tailed and Self-Similar Traffic," in *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1010–1024, 2012.

[14] M. E. Crovella, M. S. Taqqu, and A. Bestavros,. "Heavy-tailed probability distributions in the World Wide Web," in *IEEE A practical guide to heavy tails*,pp. 3-26, 1998.

[15] M. B. Møller and M. Olsen, "Sound zones: On envelope shaping of FIR filters," in *24th International Congress on Sound and Vibration, ICSV24*, pp. 613–620, 2017.