ROBUST FIR FILTERS FOR WIRELESS LOW-FREQUENCY SOUND ZONES

Mo Zhou[⁎]*, Martin Bo Møller*[†]*, Christian Sejer Pedersen*[⁎] *and Jan Østergaard*[⁎]

[⁎]Aalborg University, Aalborg, Denmark
[†]Bang & Olufsen, Struer, Denmark

ABSTRACT

Low frequency personal sound zones can be created by controlling the sound pressure in separate spatially confined regions. The performance of a sound zone system using wireless communication may be degraded due to potential packet losses. In this paper, we propose robust FIR filters for low-frequency sound zone system by incorporating information about the expected packet losses into the design. A simulation study with eight loudspeakers surrounding two control regions shows that the proposed filters can improve the contrast and the sound quality when packet losses occur, with only a slight degradation in performance even when there is no packet loss. With the proposed filters, it is possible to gain 2 dB higher contrast on average across the frequency range 20-200Hz relative to the original filters when packet loss rate is 5%.

Index Terms— Sound zone, low frequency, wireless communication, packet loss, robustness

1. INTRODUCTION

Sound zones are often considered to provide individual audio experiences to multiple listeners in the same room without the use of headphones [1]. This can be achieved by utilizing an array of loudspeakers to generate desired sound fields at predetermined locations in the room [2, 3, 4]. The control scheme often employs a set of finite impulse response (FIR) filters, which is applied for each individual loudspeaker. In order to reproduce individual audio content in multiple zones, one zone is defined as the bright zone where the sound is desired, and all the other zones are regarded as dark zones where the sound should be suppressed. Combined with the principle of superposition, sound zones with different audio content can be created. Since the wave length across the audible frequencies varies significantly, different control strategies are usually considered in different frequency ranges [5]. In this paper, we focus on the creation of sound zones at low frequencies, where the degrees of freedom in the sound field is comparable to the number of available loudspeakers [6].

In a sound zone system, the input signal is first convolved with a set of pre-determined control filters in a server, then transmitted to each loudspeaker to create the sound zones. The transmission can be conducted via cables that connect the server to all loudspeakers, which has the advantages of high speed and robustness. During the past decades, the popularity of wireless network connection has increased due to its portability, increased flexibility, and lower installation costs, which motivates us to consider a sound zone system with wireless communication in this paper. Listeners can easily set up the system and move around freely within the room without being disturbed by the cables. However, the vulnerability of wireless network often leads to bit errors and loss of packets. Small variations in the local environment caused by e.g. people walking around, can also undermine system performance significantly [7]. For sound zones, it is not only the audio quality that suffers, but especially the leakage to the dark zones is increased, resulting in lower contrast and audible artefacts [8]. Therefore, robustness to packet losses is crucial to high quality audio transmission in wireless channels where data loss often happens.

In this paper, we focus on the derivation of robust filters for wireless low-frequency sound zone systems. Based on the assumption of independent and identically distributed (i.i.d.) packet loss in each channel, a set of robust FIR filters is determined for each loudspeaker to pre-process the playback signal given the information about the state of the wireless channels. A simulation study is conducted to show that the proposed filters can improve the contrast and sound quality when packet losses occur.

2. THEORY

2.1. Preliminaries

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

$$p_{m,l}(n) = \sum_{j=0}^{J-1} h_{m,l}(j) \sum_{i=0}^{I-1} w_l(i) x_s(n-i-j), \tag{1}$$

where x_s is the input audio signal. Assuming the source signal x_s to be spectrally flat, it can be simplified as an unit sample sequence and equation (1) can be written as

$$\boldsymbol{p}_{m,l} = \boldsymbol{H}_{m,l}\boldsymbol{w}_l, \tag{2}$$

where $\boldsymbol{p}_{m,l} \in \mathbb{R}^{I+J-1}$ and $\boldsymbol{H}_{m,l} \in \mathbb{R}^{(I+J-1)\times I}$ is a Toeplitz matrix defined as:

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

2.2. Model with i.i.d. packet loss

The input signal is filtered by the FIR filters first in the server, then the filtered signal is sent to L loudspeakers by wireless communication which corresponds to L channels, where the packet loss may occur. We make the simplifying assumption that the packet size is 1. For each channel l, we assume independent packet loss, which means that each packet has probability p_l to be lost. Denote $\delta_l(t)$ as the observation of whether a packet is lost, with $P(\delta_l(t)=0)=p_l$ and $P(\delta_l(t)=1)=1-p_l$. The sound pressure under packet loss can be written as:

$$\boldsymbol{p}_{m,l} = \boldsymbol{H}_{m,l}\mathrm{diag}(\boldsymbol{\delta}_l)\boldsymbol{w}_l,$$

where $\boldsymbol{\delta}_l = (\delta_l(0), ..., \delta_l(I-1))^T$.

Let

$$\begin{aligned}&\boldsymbol{H} = (\boldsymbol{H}_1^T, ..., \boldsymbol{H}_M^T)^T \text{ with } \boldsymbol{H}_m = (\boldsymbol{H}_{m,1}, ..., \boldsymbol{H}_{m,L}),\\ &\boldsymbol{\delta} = (\boldsymbol{\delta}_1^T, ..., \boldsymbol{\delta}_L^T)^T, \boldsymbol{w} = (\boldsymbol{w}_1^T, ..., \boldsymbol{w}_L^T)^T, \Delta = \mathrm{diag}(\boldsymbol{\delta}),\end{aligned}$$

the sound pressure for M microphone positions can be written as

$$\boldsymbol{p} = \boldsymbol{H}\Delta\boldsymbol{w}. \tag{3}$$

Let $\boldsymbol{H}_B$, $\boldsymbol{H}_D$ be the RIRs for the bright zone and the dark zone respectively, since when the packet losses occur, both zones will be affected, we can write the sound pressure for the bright and dark zones as

$$\boldsymbol{p}_B = \boldsymbol{H}_B\Delta\boldsymbol{w}, \quad \boldsymbol{p}_D = \boldsymbol{H}_D\Delta\boldsymbol{w}. \tag{4}$$

2.3. Cost Function

We will use the following cost function:

$$J_{pl}(\boldsymbol{w}) = (1-\beta)\mathbb{E}\{\|\boldsymbol{p}_B - \boldsymbol{p}_T\|_2^2\} + \beta\mathbb{E}\{\|\boldsymbol{p}_D\|_2^2\} + \lambda_w \boldsymbol{w}^T\boldsymbol{R}_w\boldsymbol{w}, \tag{5}$$

where $\boldsymbol{p}_T$ is the target sound pressure in the bright zone and $\boldsymbol{R}_w$ is a weighting matrix for controlling the shape of the resulting FIR filters as suggested in [9]. The expectation $\mathbb{E}(\cdot)$ is with respect to the packet loss Δ. The FIR filters $\boldsymbol{w}$ can be estimated by minimizing (5) and the solution can be derived as

$$\begin{aligned}\boldsymbol{w}_{opt} \;=\; &[(1-\beta)\mathbb{E}(\Delta\boldsymbol{H}_B^T\boldsymbol{H}_B\Delta) + \beta\mathbb{E}(\Delta\boldsymbol{H}_D^T\boldsymbol{H}_D\Delta)\\ &+\lambda_w\boldsymbol{R}_w]^{-1}(1-\beta)\mathbb{E}(\Delta)\boldsymbol{H}_B^T\boldsymbol{p}_T.\end{aligned} \tag{6}$$

The expectation of Δ is $\mathbb{E}(\Delta) = \mathbb{E}(\mathrm{diag}(\boldsymbol{\delta})) = \Psi \otimes \boldsymbol{I}_I$, where $\boldsymbol{I}_I$ is an I-by-I identity matrix and $\Psi = \mathrm{diag}(1-p_1, ..., 1-p_L)$. Here, $\otimes$ denotes Kronecker product. For the expectation of the second moment term, each block of $\mathbb{E}(\Delta\boldsymbol{H}_B^T\boldsymbol{H}_B\Delta)$ has the form

$$\begin{aligned}&[\mathbb{E}(\Delta\boldsymbol{H}_B^T\boldsymbol{H}_B\Delta)]_{l_1,l_2}\\ =\; &\sum_{m=1}^{M}\mathbb{E}(\mathrm{diag}(\boldsymbol{\delta}_{l_1})\boldsymbol{H}_{B,m,l_1}^T\boldsymbol{H}_{B,m,l_2}\mathrm{diag}(\boldsymbol{\delta}_{l_2})),\end{aligned}$$

where $l_1, l_2 = 1, ..., L$. For given m, l_1, l_2, the expectation can be written as

$$\begin{aligned}&\mathbb{E}(\mathrm{diag}(\boldsymbol{\delta}_{l_1})\boldsymbol{H}_{B,m,l_1}^T\boldsymbol{H}_{B,m,l_2}\mathrm{diag}(\boldsymbol{\delta}_{l_2}))\\ =\; &(\boldsymbol{H}_{B,m,l_1}^T\boldsymbol{H}_{B,m,l_2}) \odot (\mathbb{E}\{\Xi_{l_1,l_2}\}),\end{aligned} \tag{7}$$

where $\odot$ denotes Hadamard product and

$$\Xi_{l_1,l_2} = \begin{pmatrix} \delta_{l_1}(0)\delta_{l_2}(0) & \cdots & \delta_{l_1}(0)\delta_{l_2}(I-1)\\ \delta_{l_1}(1)\delta_{l_2}(0) & \cdots & \delta_{l_1}(1)\delta_{l_2}(I-1)\\ \vdots & \ddots & \vdots\\ \delta_{l_1}(I-1)\delta_{l_2}(0) & \cdots & \delta_{l_1}(I-1)\delta_{l_2}(I-1)\end{pmatrix}.$$

If $l_1 = l_2 = l$, equation (7) is

$$\begin{aligned}&\mathbb{E}(\mathrm{diag}(\boldsymbol{\delta}_l)\boldsymbol{H}_{B,m,l}^T\boldsymbol{H}_{B,m,l}\mathrm{diag}(\boldsymbol{\delta}_l))\\ =\; &(1-p_l)^2\boldsymbol{H}_{B,m,l}^T\boldsymbol{H}_{B,m,l}\odot\Omega_l,\end{aligned}$$

where $\Omega_l = \mathbf{1}_I + (\frac{1}{1-p_l} - 1)\boldsymbol{I}_I$ and $\mathbf{1}_I$ is an I-by-I matrix with all elements as 1. Here, we use the fact that $\mathbb{E}\{\delta_l(t)^2\} = 1-p_l$ and $\mathbb{E}(\delta_l(t_1)\delta_l(t_2)) = (\mathbb{E}\{\delta_l(t_1)\})(\mathbb{E}\{\delta_l(t_2)\}) = (1-p_l)^2$ for $t_1 \neq t_2$.

If $l_1 \neq l_2$, equation (7) becomes

$$\begin{aligned}&\mathbb{E}(\mathrm{diag}(\boldsymbol{\delta}_{l_1})\boldsymbol{H}_{B,m,l_1}^T\boldsymbol{H}_{B,m,l_2}\mathrm{diag}(\boldsymbol{\delta}_{l_2}))\\ =\; &(1-p_{l_1})(1-p_{l_2})\boldsymbol{H}_{B,m,l_1}^T\boldsymbol{H}_{B,m,l_2}.\end{aligned}$$

Therefore, we have

$$\mathbb{E}(\Delta\boldsymbol{H}_B^T\boldsymbol{H}_B\Delta) = \boldsymbol{H}_B^T\boldsymbol{H}_B\odot\Omega,$$

where

$$\Omega = \begin{pmatrix} (1-p_1)^2\Omega_1 & \cdots & (1-p_1)(1-p_L)\mathbf{1}_I\\ (1-p_1)(1-p_2)\mathbf{1}_I & \cdots & (1-p_2)(1-p_L)\mathbf{1}_I\\ \vdots & \ddots & \vdots\\ (1-p_1)(1-p_L)\mathbf{1}_I & \cdots & (1-p_L)^2\Omega_L\end{pmatrix}.$$

Similarly, for the dark zone,

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

Thus, equation (6) has the form

$$\begin{aligned}\boldsymbol{w}_{opt} \;=\; &[(1-\beta)\boldsymbol{H}_B^T\boldsymbol{H}_B\odot\Omega + \beta\boldsymbol{H}_D^T\boldsymbol{H}_D\odot\Omega\\ &+\lambda_w\boldsymbol{R}_w]^{-1}(1-\beta)(\Psi\otimes\boldsymbol{I}_I)\boldsymbol{H}_B^T\boldsymbol{p}_T.\end{aligned} \tag{8}$$

3. SIMULATION

3.1. 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 [10], 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 1.

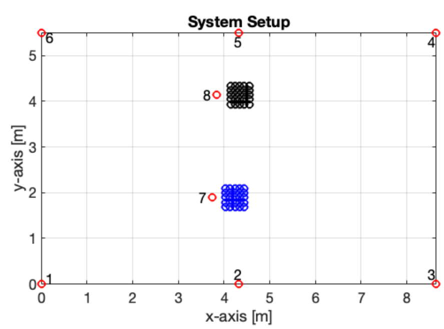


Fig. 1: System setup for the simulation. The blue and black circles are the microphones in the bright zone and dark zone respectively. The red circles are the loudspeakers.

We focus on the cases where only one channel is subject to packet loss in our simulations. Denote $\boldsymbol{\omega}_{i,p}, i = 1, ..., 8$ as our proposed filters derived by assuming Channel i has packet loss rate p, we vary the packet loss rate as $p = 5\%, 10\%, 15\%$ to see how these filters behave. The perfomances of $\boldsymbol{\omega}_{i,p}$ are compared with the original filters which are derived without packet loss, denoted by $\boldsymbol{\omega}_{old}$. 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 packet loss conditions. The weighting matrix $\boldsymbol{R}_w$ is chosen according to [9].

The transmission is simulated by using 10s input signal sampled at 1200 Hz, leading to $N = 12000$ samples in total. We encode the input signal into consecutive and non-overlapping blocks, each consists of 24 samples, which corresponds to 20 ms per block.

To evaluate the performance, Smoothed Contrast and Mean Contrast are calculated to demonstrate the performance of the sound zone system. In addition, it is known that the Perceptual Evaluation of Audio Quality (PEAQ) [11, 12, 13] is widely used to provide accurate evaluation of audio quality degradation occurring through coding procedures. We use the Objective Difference Grade (ODG) which is an output of the PEAQ model to evaluate the quality of the reproduced sound zone in the bright zone, and the reduction in the sound quality due to leaked sound from another zone[1]. The range of the ODG scales from 0 (imperceptible) to -4 (very annoying). The signals are resampled to 48 kHz before the evaluation which is the sampling frequency expected by the PEAQ model. The input signal is chosen as Gaussian white noise for the Smoothed Contrast and Mean Contrast results, and a 10-second segment of *Dazed and Confused* from the Album "Led Zeppelin" is used for the PEAQ model results.

3.2. Contrast

The Smoothed Contrast is derived by smoothing sound pressure in the frequency domain by 1/12-th Octave. Denote $\hat{p}_{B,m,l}, \hat{p}_{D,m,l}$ the reproduced sound pressures in the bright and dark zones respectively using (1) with $w_l(i)$ substituted by the estimated FIR filters, the Smoothed Contrast is calculated by

$$SC(\omega) = 10 \times \log_{10}\left(\frac{S(\sum_{m=1}^{M}(\sum_{l=1}^{L}\tilde{\hat{p}}_{B,m,l}(\omega))^2/M_B)}{S(\sum_{m=1}^{M}(\sum_{l=1}^{L}\tilde{\hat{p}}_{D,m,l}(\omega))^2)/M_D}\right),$$

where $\tilde{\hat{p}}_{B,m,l}, \tilde{\hat{p}}_{D,m,l}$ are the Fourier transform of $\hat{p}_{B,m,l}, \hat{p}_{D,m,l}$ respectively, and $S(\cdot)$ is the smoothing function.

Figure 2 plots the Smoothed Contrast of $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{5,p}$ when evaluated with and without packet loss in Channel 5 [2]. The contrasts of $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{5,p}$ are similar when no packet loss occurs. When $p = 5\%$, $\boldsymbol{\omega}_{5,p}$ gives 2 dB higher contrast on average across the frequency range 20-200 Hz relative to $\boldsymbol{\omega}_{old}$. When the packet loss rate p increases (from (a) to (c)), the contrast of $\boldsymbol{\omega}_{old}$ decreases significantly while the one of $\boldsymbol{\omega}_{5,p}$ is robust and closer to the contrast of no packet loss case.

The Mean Contrast is defined as

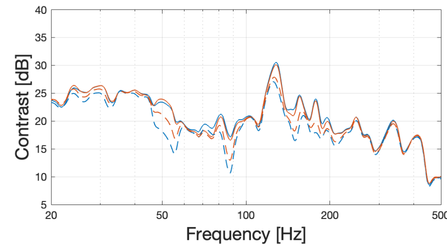
$$MC = \frac{1}{N}\sum_{n=1}^{N} C_T(n),$$

where

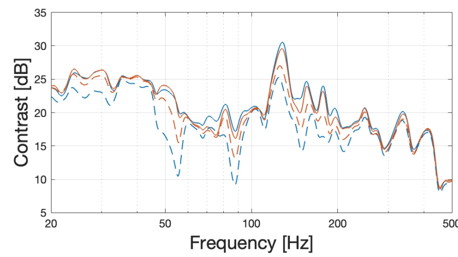
$$C_T(n) = 10 \times \log_{10}\left(\frac{\sum_{m=1}^{M}(\sum_{l=1}^{L}\hat{p}_{B,m,l}(n))^2}{\sum_{m=1}^{M}(\sum_{l=1}^{L}\hat{p}_{D,m,l}(n))^2}\right).$$

[1]We use the basic version of the implementations for the PEAQ model which is based on the Matlab code from McGill [14].

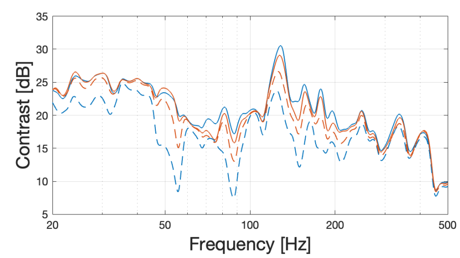
[2]Due to the limitation of space, we only present results for Channel 5 here.



(a) $p = 5\%$



(b) $p = 10\%$



(c) $p = 15\%$

Fig. 2: Smoothed contrast of $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{5,p}$ when evaluated with and without packet loss in Channel 5. (——): $\boldsymbol{\omega}_{old}$ evaluated with no packet loss. (——): $\boldsymbol{\omega}_{5,p}$ evaluated with no packet loss. (- - - -): $\boldsymbol{\omega}_{old}$ evaluated with p packet loss. (- - - -): $\boldsymbol{\omega}_{5,p}$ evaluated with p packet loss.

Figure 3 plots the Mean Contrast for $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{i,p}$ under different packet loss rates. It is clear to see that $\boldsymbol{\omega}_{i,p}$ generally has higher Mean Contrast.

3.3. Sound Quality of the Reproduced Sound in Bright Zone

In this subsection, we evaluate the quality of the reproduced audio in the bright zone for both $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{i,p}$. The reference signal for the PEAQ model is derived as the response from the reference loudspeaker (Loudspeaker 7) to the sampling positions in the bright zone. Figure 4 plots the PEAQ ODG of bright zone for $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{i,p}$ under different packet loss rates and indicates that $\boldsymbol{\omega}_{i,p}$ generally has larger ODG than $\boldsymbol{\omega}_{old}$.

3.4. Sound Quality Reduction due to Leakage

In this subsection, we evaluate the sound quality reduction of listening to the audio when exposed to leakage as well as the intended audio. To isolate the effects of the leakage, we consider a perfect intended audio reproduction, which is not subject to the sound quality reduction due to the sound zones processing nor the packet losses.

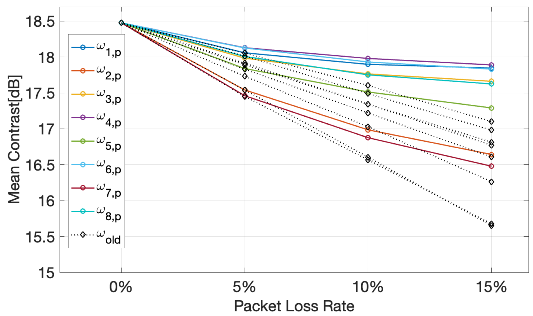


Fig. 3: Mean Contrast for $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{i,p}$ under different packet loss rates. Solid lines represent Mean Contrast when evaluated with p packet loss in Channel i using $\boldsymbol{\omega}_{i,p}$. ····◇··· represents Mean Contrast when evaluated with p packet loss in Channel i using $\boldsymbol{\omega}_{old}$.

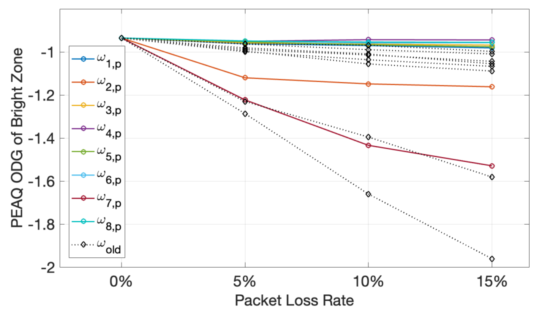


Fig. 4: PEAQ ODG of bright zone for $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{i,p}$ under different packet loss rates.

More specifically, imagine that we have two zones, denoted as Zone A and Zone B. The wireless sound zone system reproduces audio signal A in Zone A and seeks to reduce the leakage towards Zone B. In addition, we have a loudspeaker (unaffected by wireless losses or sound zones processing) reproducing audio signal B for Zone B. We then evaluate how a person in Zone B experiences the quality of listening to the combination of the reproduced audio signal B and the leaked audio from Zone A under different packet loss patterns.

We treat the bright zone in previous section as Zone A and choose the input signal for Zone B as a 10-second segment of *Whole Lotta Love* from the Album "Led Zeppelin II". A calibration procedure is conducted to ensure that the reproduced audio has equal loudness[3] in the two zones. Figure 5 plots the PEAQ ODG of Zone B for $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{i,p}$ under different packet loss rates and shows that our proposed filters $\boldsymbol{\omega}_{i,p}$ leads to a marginally lower reduction in the sound quality in Zone B due to leakage from Zone A.

4. DISCUSSION

From the results presented in the previous section, it is clearly seen that incorporating packet loss information can improve the performance of the wireless low-frequency sound zone system. The contrast performance can be seen from the comparison of $\boldsymbol{\omega}_{i,p}$ plotted against $\boldsymbol{\omega}_{old}$ in Figure 2-3. It is shown that our proposed filters $\boldsymbol{\omega}_{i,p}$ not only improves the overall contrast when packet losses occur, but also has comparable performance to $\boldsymbol{\omega}_{old}$ when evaluated with no packet loss. When p increases, the improvement is more siginificant for frequency range 20-200 Hz, indicating the robustness of our proposed filters. Furthermore, the evaluation procedure using the PEAQ model shows that packet loss indeed leads to a lower sound quality in the bright zone (Figure 4). In Figure 5, the differences among the ODG values are almost negligible and all the values are close to 0 which indicates that the changes in the sound quality are imperceptible. The undesired leakage under packet loss has only minor influence on listener's experience in another zone. This is likely due to the reason that we can attain a high contrast in general in this scenario, where the reproduced audio in Zone B is free of streaming artifacts. The new filters $\boldsymbol{\omega}_{i,p}$ can reduce the sound quality degradation in the bright zone to some degree, and at the same time have fair controllability of the leakage.

The robustness of $\boldsymbol{\omega}_{i,p}$ comes from the functionability of matrices Ω and Ψ in equation (8). When there is no packet loss, Ω reduces to a matrix with all entries as 1 and Ψ reduces to an identity matrix, then equation (8) gives result as $\boldsymbol{\omega}_{old}$. When Channel l has packet loss, matrix Ω will reduce the magnitude of the filter for Channel l and make the other channels compensate for this reduction, leading to the robust performance.

The derivation of the robust filters is based on the assumption that the packet loss channel and rate are known. In practice, it is possible to know which channel is subject to packet loss and estimate its packet loss rate by analyzing the sent and the received packets. Futhermore, the results presented in this paper are limited to the i.i.d. packet loss assumption which may be too strict in practice, hence, it would be interesting to further incorporate dependent or bursty packet loss patterns.

5. CONCLUSION

In this paper, it has been shown that robust FIR filters can be derived for wireless low-frequency sound zone system by incorporating information about the expected packet losses. The proposed robust FIR filters can improve contrast and sound quality when packet losses occur and still provide a comparable performance even when there is no packet loss. Further investigation will consider incorporating bursty packet loss and also the case when all channels have packet loss.

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).

[3]We use Matlab function `acousticLoudness`, which measures the ISO 532-1 stationary free-field loudness.

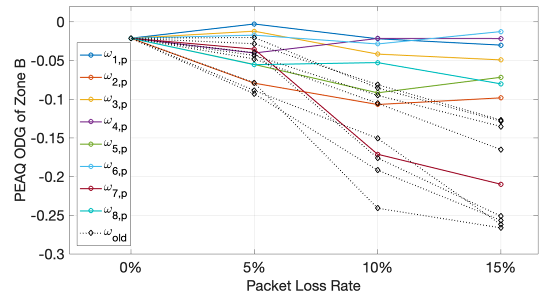


Fig. 5: PEAQ ODG of Zone B for $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{i,p}$ under different packet loss rates.

6. REFERENCES

[1] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal soundzones: Delivering interface-free audio to multiple listeners," *IEEE Signal ProcessingMagazine*, vol. 32, no. 2, pp. 81–91, March 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, April 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, April 2014.

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

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

[6] L. Brännmark, A. Bahne, and A. Ahlén, "Compensation of loudspeaker-roomresponses in a robust mimo control framework," *IEEE Transactions on Audio,Speech, and Language Processing*, vol. 21, no. 6, pp. 1201–1216, October 2013.

[7] W. Simpson, "Chapter 7: Packet transport," in *Video over IP: a practical guide to technology and applications*. 2006, Taylor & Francis.

[8] 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, May 2022.

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

[10] F. Jacobsen and P. M. Juul, *Fundamentals of General Linear Acoustics*, John Wiley & Sons, 2013.

[11] ITU-R Rec. BS.1387, *Method for objective measurements of perceived audio quality*, 2001.

[12] C. Colomes, C. Schmidmer, T. Thiede, and W.C. Treurniet, "Perceptual quality assessment for digital audio: Peaq-the new itu standard for objective measurement of the perceived audio quality," *Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding*, September 1999.

[13] M. Torcoli, T. Kastner, and J. Herre, "Objective measures of perceptual audio quality reviewed: An evaluation of their application domain dependence," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1530–1541, September 2021.

[14] P. Kabal, "An examination and interpretation of itu-r bs.1387: Perceptual evaluation of audio quality," in *TSP Lab Technical Report*. December 2003, Department of Electrical & Computer Engineering, McGill University.