Robust FIR Filters for Wireless Low-Frequency Sound Zones

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

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

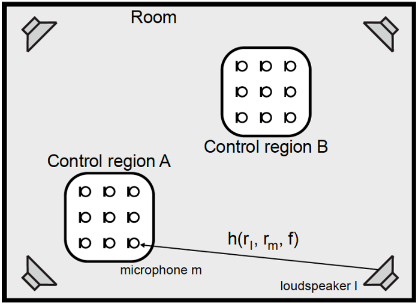
2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023)
4-10 June, Rhodes Island,Greece

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)

Outline

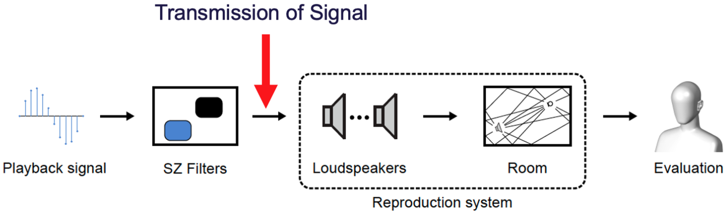
- Introduction of Sound Zone Systems
- Model with Packet Loss
- Filter Derivation
- Simulation
    - Simulation Setting
    - Contrast
    - Sound Quality
- Discussion

Sound Zone



- Bright zone: where the sound is desired
- Dark zones: where the sound is suppressed
- Different control strategies are usually considered in different frequency ranges
- we focus on the creation of sound zones at low frequencies

Sound Zone Systems



- Cable: high speed, robustness
- Wireless:
  - portability, increased flexibility, lower installation costs
  - bit errors and loss of packets

Effect of Packet Loss on Sound Zones

- Lower contrast
- Lower audio quality
- More leakage to the dark zone

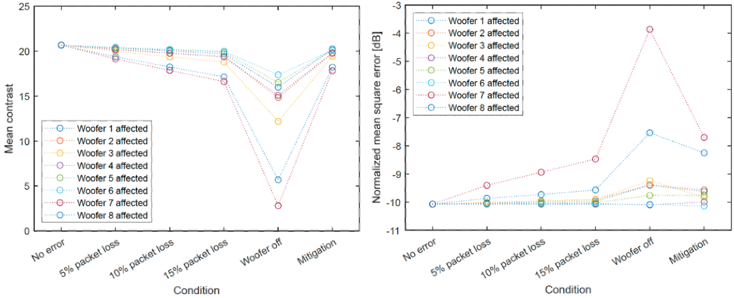


Figure: Contrast and normalized mean square error for 10s gaussian white noise depending on error and affected woofers.[1]

---

[1] 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.

Preliminaries

- The sound pressure at time n recorded by microphone m due to loudspeaker l:

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

  - x_s is the input audio signal,
  - $\boldsymbol{h}_{m,l} = (h_{m,l}(0), ..., h_{m,l}(J-1))^T$ is the room impulse response (RIR),
  - $\boldsymbol{w}_l = (w_l(0), ..., w_l(I-1))^T$ is the FIR filters.

- Assume the source signal x_s to be spectrally flat:

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

  - $\boldsymbol{H}_{m,l} \in \mathbb{R}^{(I+J-1)\times I}$ is a Toeplitz matrix.

Model with Packet Loss

- Assume the packet size is 1 and packet loss in each channel l is independent:

$$P(\delta_l(t)=0)=p_l, P(\delta_l(t)=1)=1-p_l,$$

where $\delta_l(t)$ is the observation of whether a packet is lost.

- Sound Pressure under packet loss can be written as:

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

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

- For M microphone positions:

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

where

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

Cost Function

- Sound Pressure in the bright zone and dark zone:

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

- We will use the following cost function:

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

  - $\boldsymbol{p}_T$ is the target sound pressure in the bright zone
  - $\boldsymbol{R}_w$ is a weighting matrix for controlling the shape of the FIR filters
  - The expectation $\mathbb{E}(\cdot)$ is with respect to the packet loss Δ.

Filter Derivation

- The FIR filters $\boldsymbol{w}$ can be estimated by minimizing (5):

$$\begin{aligned}\boldsymbol{w}_{opt} &= [(1-\beta)\mathbb{E}(\Delta \boldsymbol{H}_B^T \boldsymbol{H}_B \Delta) + \beta \mathbb{E}(\Delta \boldsymbol{H}_D^T \boldsymbol{H}_D \Delta) \\ &\quad + \lambda_w \boldsymbol{R}_w]^{-1} (1-\beta)\mathbb{E}(\Delta) \boldsymbol{H}_B^T \boldsymbol{p}_T. \\ &= [(1-\beta)\boldsymbol{H}_B^T \boldsymbol{H}_B \odot \Omega + \beta \boldsymbol{H}_D^T \boldsymbol{H}_D \odot \Omega \\ &\quad + \lambda_w \boldsymbol{R}_w]^{-1} (1-\beta)(\Psi \otimes \boldsymbol{I}_I) \boldsymbol{H}_B^T \boldsymbol{p}_T. \qquad (6)\end{aligned}$$

where

- $\mathbb{E}(\Delta) = \Psi \otimes \boldsymbol{I}_I$, where $\boldsymbol{I}_I$ is an I-by-I identity matrix and $\Psi = \text{diag}(1-p_1, ..., 1-p_L)$, $\otimes$ denotes Kronecker product, and

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

Simulation Setting

- 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 $M_B = M_D = 75$.

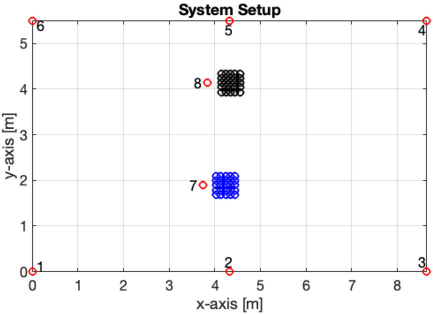


Figure: The blue and black circles are the microphones in the bright zone and dark zone respectively. The red circles are the loudspeakers.

Simulation Setting

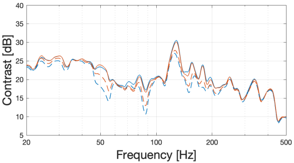
- focus cases where only one channel is subject to packet loss,
- $\boldsymbol{\omega}_{i,p}, i = 1, ..., 8$: proposed filters derived by assuming Channel i has packet loss rate p, where $p = 5\%, 10\%, 15\%$,
- $\boldsymbol{\omega}_{old}$: original filters derived without packet loss,
- The RIRs and the filters are of length $J = 600$ and $I = 300$,
- $\beta = 0.97$ and $\lambda_w = 10^{-7}$,
- Input signal: 10s signal sampled at 1200 Hz, encoded into consecutive and non-overlapping blocks, each consists of 24 samples (20 ms).

Smoothed Contrast

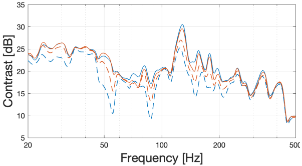
- Input signal: Gaussian white noise
- Smoothed Contrast:

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

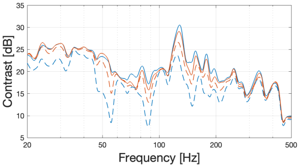
where $\tilde{\hat{p}}_{B,m,l}, \tilde{\hat{p}}_{D,m,l}$ are the Fourier transform of $\hat{p}_{B,m,l}, \hat{p}_{D,m,l}$ respectively, and $S(\cdot)$ is the smoothing function.



(a) $p = 5\%$



(b) $p = 10\%$



(c) $p = 15\%$

Mean Contrast

- Input signal: Gaussian white noise
- The Mean Contrast is defined as

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

where

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

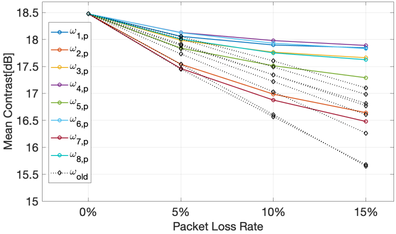


Figure: Mean Contrast for ω_{old} and $\omega_{i,p}$ under different packet loss rates.

Sound Quality in Bright Zone

- Input signal: 10-second segment of *Dazed and Confused* from the Album "Led Zeppelin" .
- We use the Perceptual Evaluation of Audio Quality (PEAQ) model to predict sound quality in the bright zone.

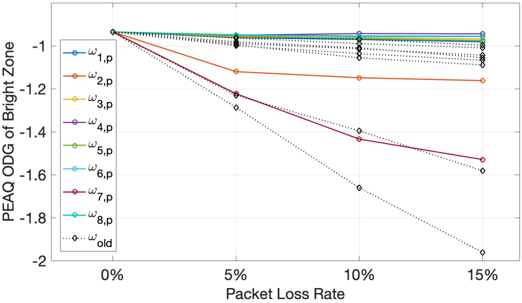


Figure: PEAQ ODG of bright zone for $\boldsymbol{\omega}_{old}$ and $\boldsymbol{\omega}_{i,p}$ under different packet loss rates.

Sound Quality Reduction due to Leakage

- We evaluate the sound quality reduction of listening to the audio when exposed to leakage as well as the intended audio.
- Assume that we have two zones (Zone A and Zone B), we reproduces audio signal A in Zone A and seeks to reduce the leakage towards Zone B. Zone B has a loudspeaker 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.

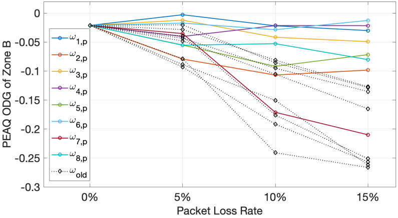


Figure: PEAQ ODG of Zone B for ω_{old} and $\omega_{i,p}$ under different packet loss rates.

Discussion

- Incorporating packet loss information can improve the performance of the wireless low-frequency sound zone system. With packet loss rate increases, the improvement is more significant,
- Our proposed filters not only improves the overall contrast when packet losses occur, but also has comparable performance when evaluated with no packet loss,
- Future investigation: incorporating bursty packet loss, cases when all channels have packet loss.

Thank you!