A UNIFIED APPROACH TO GENERATING SOUND ZONES USING VARIABLE SPAN LINEAR FILTERS

AUDIO ANALYSIS LAB

Taewoong Lee, Jesper Kjær Nielsen, Jesper Rindom Jensen, and Mads Græsbøll Christensen
Audio Analysis Lab, CREATE, Aalborg University, Denmark {tlee,jkn,jrj,mgc}@create.aau.dk

AALBORG UNIVERSITY DENMARK

Introduction

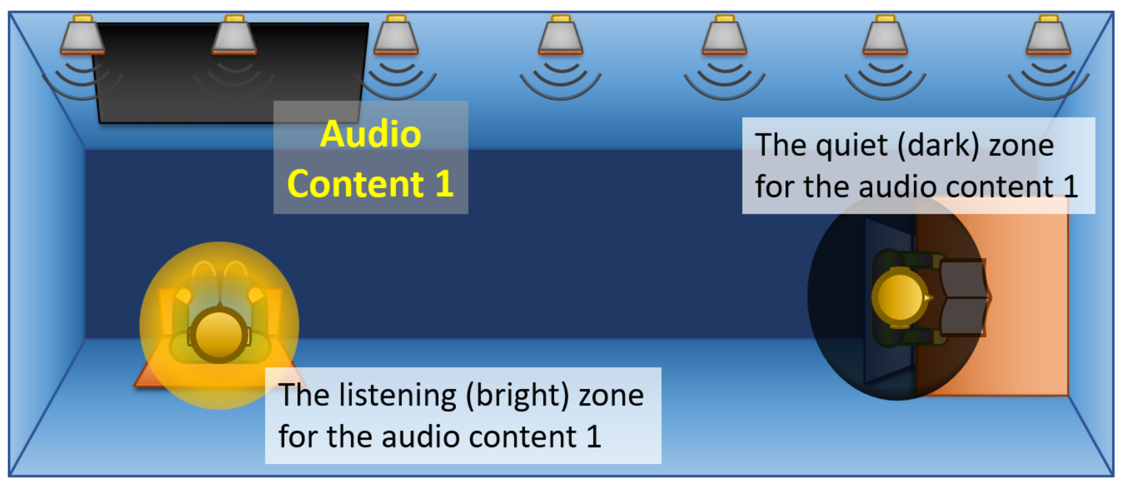

Fig. 1: An illustration of sound zones.

- Sound zones enable multiple people to enjoy different audio contents in the same acoustic space without disturbing each other.
- We propose a general framework to create sound zones. It trades-off between acoustic contrast and signal distortion.

Generation of sound zones

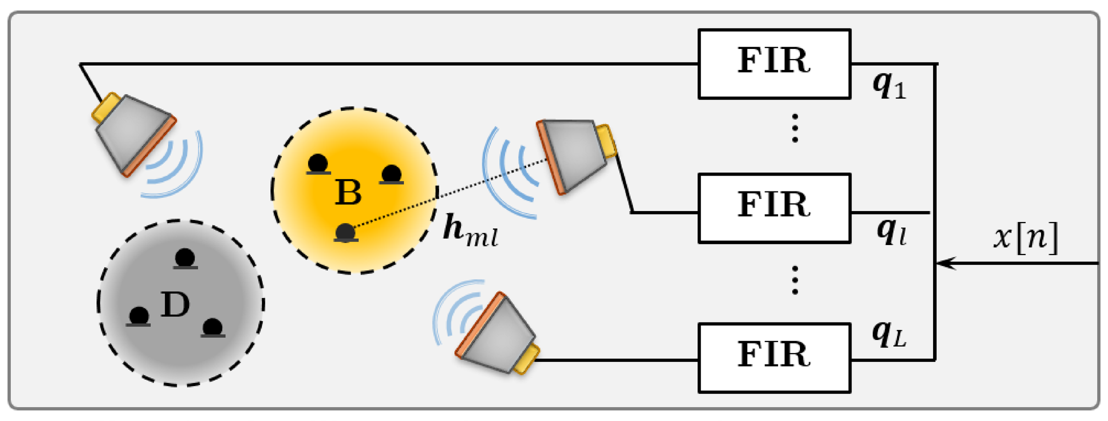

Fig. 2: An illustration of sound zones system.

The reproduced sound pressure on the mth mic. position by the L loudspeakers is

$$p_m[n] = \sum_{l=1}^{L} h_{ml}[n] * q_l[n] * x[n] = \mathbf{h}_m^T \mathbb{X}[n]\mathbf{q}, \quad (1)$$

where $h_{ml}[n]$ is the room impulse response, $\mathbf{h}_{ml} = [h_{ml}[0], \cdots, h_{ml}[K-1]]^T$, $\mathbf{h}_m = [\mathbf{h}_{m1}^T, \cdots, \mathbf{h}_{mL}^T]^T$, $\{q_l[n]\}_{l=1}^{L}$ is the control filters, $\mathbf{q}_l = [q_l[0], \cdots, q_l[J-1]]^T$, $\mathbf{q} = [\mathbf{q}_1^T, \cdots, \mathbf{q}_L^T]^T$, $\mathbf{X}[n] = \{x[n-k-j+2]\}_{k=1,j=1}^{K,J}$, $\mathbb{X}[n] = \mathbf{I}_L \otimes \mathbf{X}[n]$, and $\mathbf{I}_L = \text{diag}(1, \cdots, 1) \in \mathbb{R}^{L \times L}$.

The reproduced sound field in the bright zone $\mathbf{p}_B[n] = [p_1[n], \cdots, p_{M_B}[n]]^T$ is then

$$\mathbf{p}_B[n] = \mathbf{H}_B^T[n]\mathbf{q}, \quad (2)$$

where $\mathbf{H}_B[n] = \mathbb{X}^T[n][\mathbf{h}_1, \cdots, \mathbf{h}_{M_B}]$, and similarly $\mathbf{p}_D[n]$, $\mathbf{H}_D[n]$ for the dark zone.
For the total zone, $\mathbf{p}_C[n] = [\mathbf{p}_B^T[n], \mathbf{p}_D^T[n]]^T$, $\mathbf{H}_C[n] = [\mathbf{H}_B[n], \mathbf{H}_D[n]]$.

The desired sound field: $\mathbf{d}_C[n] = [\mathbf{d}_B^T[n], \mathbf{0}_{M_D}^T]^T$
The reproduction error : $\varepsilon_C[n] = \mathbf{d}_C[n] - \mathbf{p}_C[n]$

The acoustic potential energy:

$$e_B = \frac{1}{M_B N} \sum_{n=0}^{N-1} \mathbf{p}_B^T[n]\mathbf{p}_B[n] = \frac{1}{M_B} \mathbf{q}^T \mathbf{R}_B \mathbf{q}, \quad (3)$$

where N is the number of time samples, $\mathbf{R}_B$ is a spatial correlation matrix, and similarly e_D, $\mathbf{R}_D$ for the dark zone.
The acoustic contrast : $\gamma = e_B / e_D$

The average reproduction error energy:

$$S_C = \frac{1}{N} \sum_{n=0}^{N-1} \|\varepsilon_C[n]\|^2 = S_B + S_D, \quad (4)$$

where $\|\cdot\|$ is the ℓ_2 norm, S_B, S_D are the distortion and residual energies, respectively.

A unified framework approach

The unified framework approach is inspired by the VSLF framework [1].

- **Joint diagonalization**
  The solution to $\mathbf{R}_B\mathbf{q} = \lambda \mathbf{R}_D \mathbf{q}$ is a diagonal matrix $\mathbf{\Lambda}_{LJ}$ of the LJ eigenvalues in $\lambda_1 \geq \cdots \geq \lambda_{LJ}$, and the square matrix $\mathbf{U}_{LJ} = [\mathbf{u}_1, \cdots, \mathbf{u}_{LJ}]$ of the LJ eigenvectors.
  These jointly diagonalize $\mathbf{R}_B, \mathbf{R}_D$

$$\mathbf{U}_{LJ}^T \mathbf{R}_B \mathbf{U}_{LJ} = \mathbf{\Lambda}_{LJ}, \quad \mathbf{U}_{LJ}^T \mathbf{R}_D \mathbf{U}_{LJ} = \mathbf{I}_{LJ}. \quad (5)$$

- **Low-rank approximation**
  $\mathbf{q}$ can be written as a linear combination, i.e.,

$$\mathbf{q} = \mathbf{U}_V \mathbf{a}_V, \quad (6)$$

  where $\mathbf{U}_V$, $\mathbf{a}_V$ are the basis function and the weights, respectively. The first $1 \leq V \leq LJ$ eigenvectors in $\mathbf{U}_{LJ}$ are used as the basis function; hence, the optimization is carried out with respect to $\mathbf{a}_V$. The rank V is a user parameter which controls the trade-off between S_B and γ.

- **Variable Span Trade-off (VAST)**

$$\text{minimize } S_B \text{ subject to } S_D \leq \epsilon, \quad (7)$$

  for $1 \leq V \leq LJ$, the solution is referred to as **a variable span trade-off (VAST) filter**,

$$\mathbf{q}_{\text{VAST}}(V, \mu) = \mathbf{U}_V \mathbf{a}_V(\mu) = \sum_{v=1}^{V} \frac{\mathbf{u}_v \mathbf{u}_v^T}{\mu + \lambda_v} \mathbf{r}_B, \quad (8)$$

  where μ is the Lagrange multiplier, and $\mathbf{r}_B = N^{-1} \sum_{n=0}^{N-1} \mathbf{H}_B[n]\mathbf{d}_B[n]$. **The solution can vary with respect to** V **and** μ.

Special cases

- For $V = 1$:

$$\mathbf{q}_{\text{VAST}}(1, \mu) = \frac{\mathbf{u}_1 \mathbf{u}_1^T}{\mu + \lambda_1} \mathbf{r}_B \propto \mathbf{u}_1, \quad (9)$$

  which is the ACC [2] solution since $\mathbf{u}_1$ is the eigenvector corresponding to the largest eigenvalue.

- For $V = LJ$, $\mu = 1$:

$$\mathbf{q}_{\text{VAST}}(LJ, 1) = \mathbf{U}_V \underbrace{(\mathbf{\Lambda}_V + \mu \mathbf{I}_V)^{-1} \mathbf{U}_V^T \mathbf{r}_B}_{\mathbf{a}_V(\mu)} = (\mathbf{R}_B + \mathbf{R}_D)^{-1} \mathbf{r}_B, \quad (10)$$

  which is the PM [3] solution.

Table 1: Various solutions for sound zone control

μ	V	Analytic form
-	-	$\mathbf{q}_{\text{VAST}} = \sum_{v=1}^{V} [(\mu + \lambda_v)^{-1} \mathbf{u}_v \mathbf{u}_v^T \mathbf{r}_B]$
-	1	$\mathbf{q}_{\text{ACC}} = (\mu + \lambda_1)^{-1} \mathbf{u}_1 \mathbf{u}_1^T \mathbf{r}_B$
0	-	$\mathbf{q}_{\text{MD}} = \sum_{v=1}^{V} [\lambda_v^{-1} \mathbf{u}_v \mathbf{u}_v^T \mathbf{r}_B]$
0	LJ	$\mathbf{q}_{\text{MVDR}} = \mathbf{R}_B^{-1} \mathbf{r}_B$
1	-	$\mathbf{q}_{\text{VS W}} = \sum_{v=1}^{V} [(1 + \lambda_v)^{-1} \mathbf{u}_v \mathbf{u}_v^T \mathbf{r}_B]$
1	LJ	$\mathbf{q}_{\text{PM}} = (\mathbf{R}_B + \mathbf{R}_D)^{-1} \mathbf{r}_B$

Evaluation

- **Acoustic contrast** γ:

$$\gamma_V(\mu) = \kappa^2 \frac{\mathbf{a}_V^T(\mu) \mathbf{\Lambda}_V \mathbf{a}_V(\mu)}{\mathbf{a}_V^T(\mu) \mathbf{a}_V(\mu)} \quad (11)$$

  leads to $\gamma_V(\mu) \geq \gamma_{V'}(\mu)$ if $V \leq V'$.

- **Signal distortion** S_B:
  $S_C(V) = S_B(V) + S_D(V)$ as a function of V

$$S_B(V) = \sigma_d^2 - \sum_{v=1}^{V} \frac{\lambda_v + 2\mu}{(\lambda_v + \mu)^2} \|\mathbf{u}_v^T \mathbf{r}_B\|^2, \quad (12)$$

$$S_D(V) = \sum_{v=1}^{V} \frac{1}{(\lambda_v + \mu)^2} \|\mathbf{u}_v^T \mathbf{r}_B\|^2, \quad (13)$$

  where $\sigma_d^2 = N^{-1} \sum_{n=0}^{N-1} \|\mathbf{d}_B[n]\|^2$.

Proof-of-concept simulation

- 1 m radius circular array with three loudspeakers, three microphone positions in each zone, free field, ideal loudspeaker and microphone.
- The length of $\{\mathbf{q}_l\}_{l=1}^{3}$ is 50 ms, f_s = 12.8 kHz. $\mathbf{d}_B[n] = \mathbf{H}_B^T[n]\mathbf{i}_J^{(L)}$, where $\mathbf{i}_J^{(L)} = \mathbf{1}_L \otimes \mathbf{i}_J$, $\mathbf{1}_L = [1, \cdots, 1]^T \in \mathbb{R}^{L \times 1}$, $\mathbf{i}_J$ is the first column of $\mathbf{I}_J$, $\mu = 1$, and $x[n] = \delta[n]$.
- γ w.r.t. frequency, V, and S_B w.r.t. V.

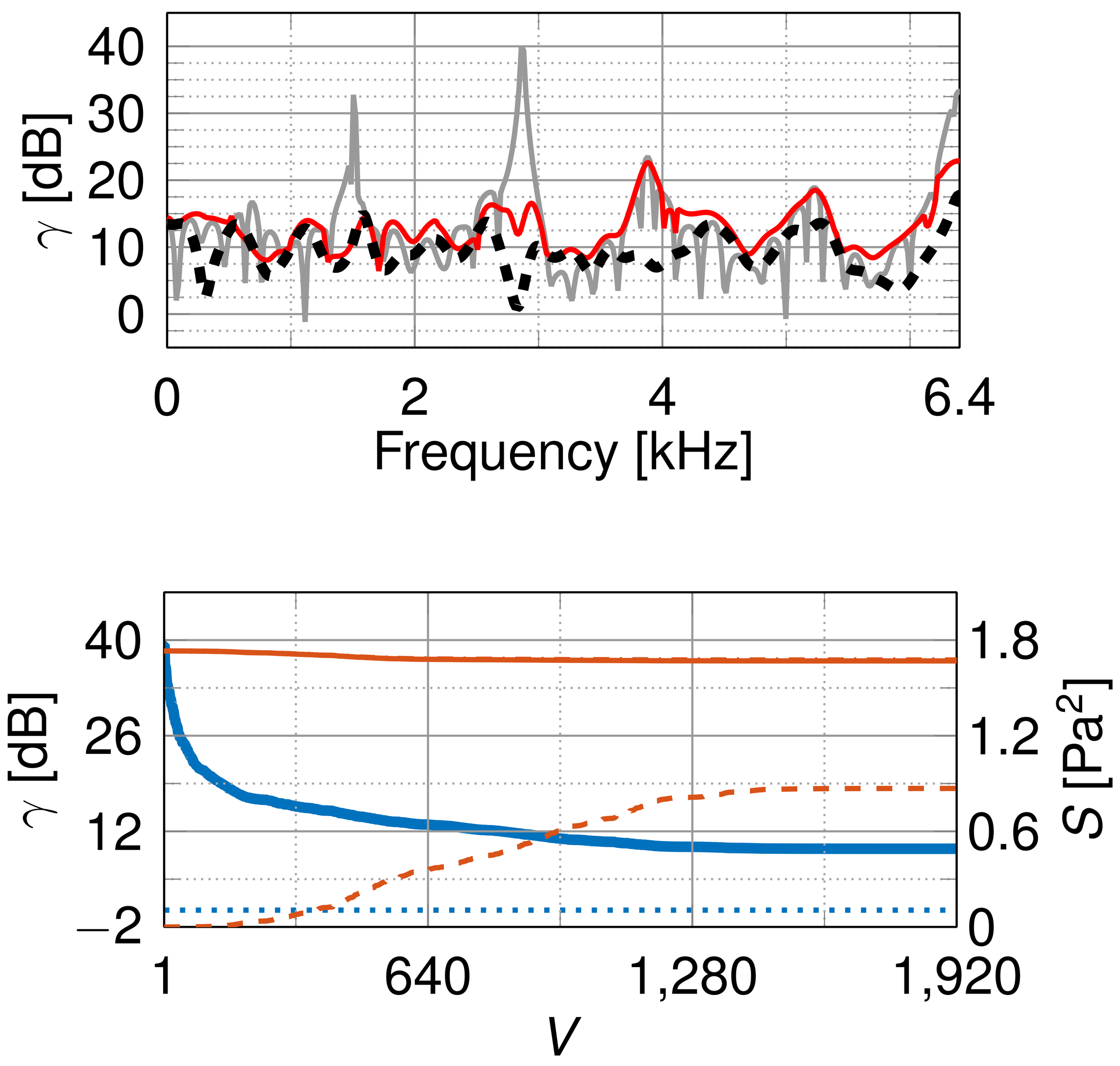

Fig. 3: $\gamma(f)$ when $V = 1$ which corresponds to ACC(—), V = 800 (—), and V = 1920 which corresponds to PM(▪▪), $\gamma(V)$ before filtering(···), $\gamma(V)$ after filtering(—), S_B(—), S_D(‑‑), and S_C(‑‑). Note that S_D is multiplied by 150 for better visualization.

Conclusion

- A new framework to create sound zones.
- ACC and PM as special cases.
- Performance evaluation based on γ and S_B.
- All solutions will have γ and S_B with upper and lower bounded by ACC and PM.

References

[1] J. R. Jensen, J. Benesty, and M. G. Christensen, “Noise reduction with optimal variable span linear filters,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 631–644, Apr. 2016.

[2] J.-W. Choi and Y.-H. Kim, “Generation of an acoustically bright zone with an illuminated region using multiple sources,” *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700, 2002.

[3] M. A. Poletti, “An investigation of 2d multizone surround sound systems,” in *Proc. 125th Conv. Audio. Eng. Soc.*, San Francisco, USA, 2008.