Perceptually Optimized Personal Sound Zones

Tanay Mannikar

EE 395 Final Project

Introduction

Personal Sound Zones (PSZs)

- Localized sound field control (dereverberation)
- Loudspeaker arrays generate isolated audio programs in two individual zones via PSZ filtering

Poses a complex physical optimization problem - adaptive filtering to the rescue!

- Define optimal filtering solution
- Facilitate real-time implementation

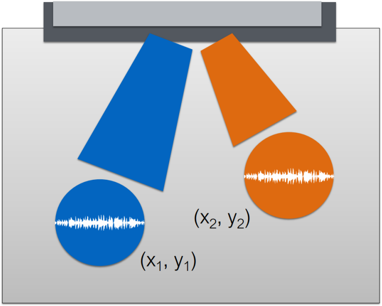
Perceptual optimization

- Account for psychoacoustic phenomena
- Enhance subjective performance measures

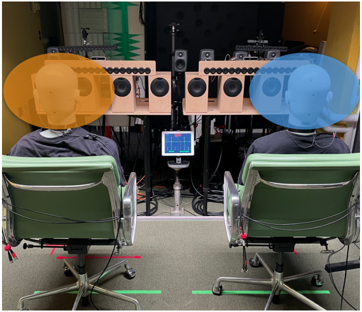
Personal Sound Zones

Applications

- Automotive cabins
- Individualized content by language or preference



(a) Bright and dark zones

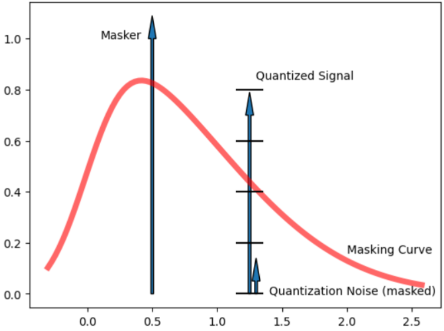


(b) Laboratory PSZ system

The *13 dB Miracle*

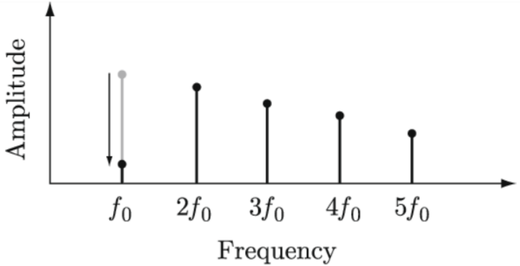
Psychoacoustic masking

- Backbone of audio compression and perceptual audio coding, i.e. MPEG-3
- Louder frequencies mask surrounding information - reproduction and quantization errors within masked regions are perceptually nonexistent
- 13 dB of masked noise nearly imperceptible!



Demo: white vs shaped noise

Missing-Fundamental Phenomenon



Pitch Perception

- Overtones or *harmonics* contribute to perceived pitch
- Removing $f_0 \rightarrow$ **virtual pitch**
- Nearly indistinguishable from pitch containing f_0!

Demo: true vs virtual pitch of a sawtooth wave, $f_0 = \{262, 440\}$ Hz

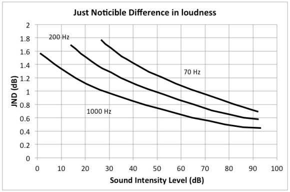
Just-Noticeable Difference (JND)

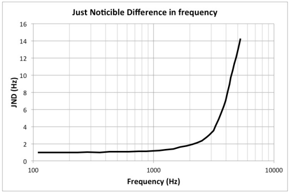
Monaural

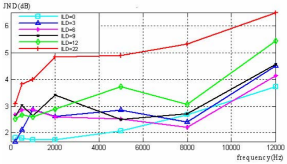
- JND-dB
- JND-Hz

Binaural

- JND-ILD (interaural level difference)
- JND-ITD (interaural time difference)





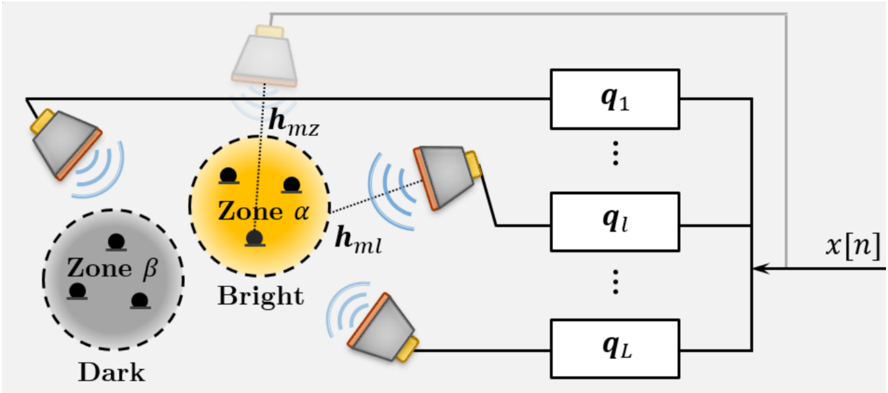


Static PSZ Formulation

Zone control point m, time index n, L loudspeakers, J FIR coefficients

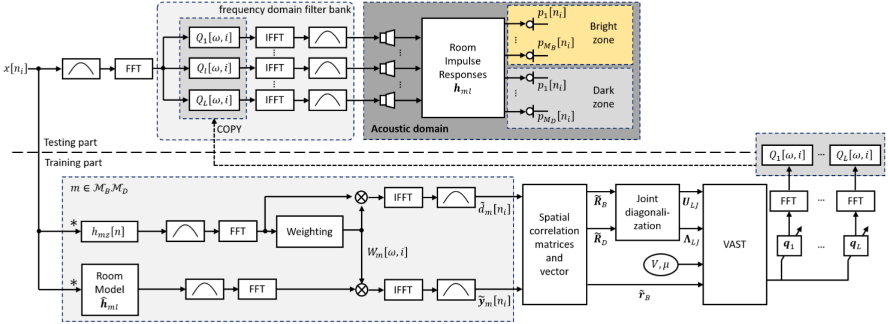
Input signal	$x[n]$
Room impulse responses (RIRs)	$\boldsymbol{h}_{ml}$
Uncontrolled pressure	$\boldsymbol{y}_{ml}[n] = \boldsymbol{X}[n]\boldsymbol{h}_{ml}$
Optimal $L \times J$ FIR PSZ filter matrix	$\boldsymbol{q}$
Optimal pressure	$p_m[n] = \sum_{l=1}^{L} \boldsymbol{y}_{ml}^T[n]\boldsymbol{q}_l = \boldsymbol{y}_m^T[n]\boldsymbol{q}$
Desired signal	$d_m[n] = \begin{cases} (h_{mz} * x)[n], & BZ \\ 0, & DZ \end{cases}$
Error (VAST)	$\varepsilon_m[n] = d_m[n] - p_m[n]$
Error (AP-VAST)	$(\varepsilon_m * w_m)[n] = \tilde{\varepsilon}_m[n] = \tilde{d}_m[n] - \tilde{p}_m[n]$

where $w_m[n]$ is the time-varying inverse of the psychoacoustic masking filter at point m and time n!


$\boldsymbol{q}_1$
$\boldsymbol{q}_l$
$\boldsymbol{q}_L$
$\boldsymbol{h}_{mz}$
$\boldsymbol{h}_{ml}$
$x[n]$
Zone α
Bright
Zone β
Dark

AP-VAST

Adaptive and perceptually optimized variable-span tradeoff (AP-VAST) filtering

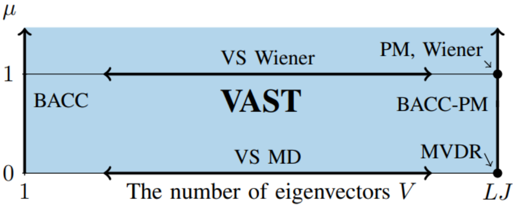


Optimal Solution

- Sparing you the details...

$$\boldsymbol{q}_o(V,\mu) = \boldsymbol{U}_V \boldsymbol{a}_o(V,\mu) = \sum_{v=1}^{V} \frac{\boldsymbol{u}_v^T \tilde{\boldsymbol{r}}_B}{\lambda_v + \mu} \boldsymbol{u}_v, \quad 1 \leq V \leq LJ$$

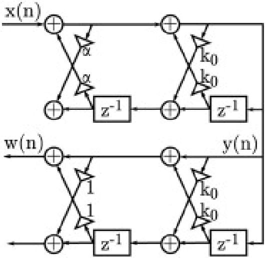
- Manipulating number of eigenvalues/vectors V and Lagrange multiplier μ provides many other existing adaptive solutions
- AP-VAST provides a 20% increase in subjective performance compared to the standard VAST approach!
- **Demo:** Pressure-matching (PM) vs AP-VAST



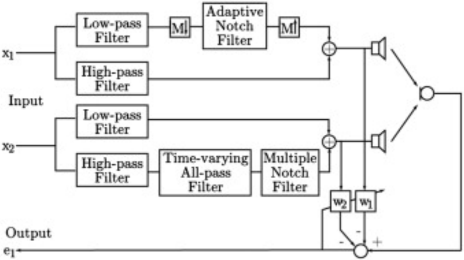
Adaptive Multichannel Decorrelation

Comb filtering effect due to linear loudspeaker array geometry causes audible distortion

- Solution: decorrelate adjacent multichannel signals by adaptive notch filtering and time-varying all-pass filtering



(a) Adaptive 2nd-order IIR notch filter



(b) Decorrelation algorithm block diagram

1) 2nd-order f_0-tracking lattice notch filter:

$$H(z) = \frac{W(z)}{X(z)} = \frac{1 + 2k_0[n]z^{-1} + z^{-2}}{1 + k_o[n](1+\alpha)z^{-1} + \alpha z^{-2}}, \quad W(z) = \sum_{k=-\infty}^{\infty} w[k]z^{-k} \tag{1}$$

$$k_0[n] = \frac{2}{1 + e^{-\mathbf{g_0}[n]}} \tag{2}$$

$$\nabla_{\mathbf{g_0}} \left(\sum_{k=0}^{n} \lambda^{n-k} w^2[k] \right) = 0, \quad 0 < \lambda < 1 \tag{3}$$

$$\hat{f}_0[n] = \frac{f_s}{2\pi M} \cos^{-1}(-k_0[n]) \tag{4}$$

2) Time-varying 2nd-order all-pass filter:

$$A(z) = \frac{k_0^2[n] - 2k_0[n]z^{-1} - z^{-2}}{1 - 2k_0[n]z^{-1} + k_0^2[n]z^{-2}}$$

- Perturbs group delay across frequency range within JND-ITD threshold ($< 40\ \mu$s)
- Preserves magnitude response (all-pass)
- Imperceptible phase distortion decorrelates adjacent channels!

3) Adaptive polynomial multiple-notch filter:

$$M(z) = \frac{\prod_{m=M_{min}}^{M_{max}} (1 - e^{j\omega_m(n)} z^{-1})}{\prod_{m=M_{min}}^{M_{max}} (1 - e^{j\omega_m(n)} \rho z^{-1})}$$

$$\omega_m(n) = 2\pi m \hat{f}_0[n], \quad 1 \leq M_{min} < m < M_{max} \leq \left\lfloor \frac{f_s}{\hat{f}_0[n]} \right\rfloor$$

- Create the lowest frequency notch within the high frequency range due to perceptual insensitivity of phase and magnitude distortion at increasingly higher frequencies

Decorrelation Performance (MSC)

Between two channels x_i and x_j, define the Magnitude Square Coherence (MSC):

$$\mathrm{MSC}_{i,j}(f) = \frac{|S_{ij}(f)|^2}{S_{ii}(f)S_{jj}(f)}$$

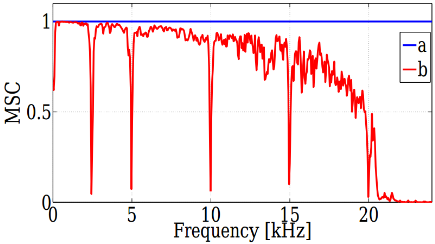


Figure: Magnitude square coherence before (a) and after (b) adaptive decorrelation method for $f_0 = 2.5$ kHz and harmonics 5 kHz, 10 kHz, 15 kHz, and 20 kHz

Conclusions

- Background, motivation, and formulation of perceptually motivated adaptive filtering algorithms for PSZs
- Possible to improve on the subjective performance of existing filtering techniques without only considering optimization of physical parameters
- Able to efficiently improve performance through imperceptible filtering operations
- Techniques can be integrated within deep neural network (DNN) methods for PSZ filter generation
- Extends to scenarios involving head-tracking to further improve listening performance of moving zones and head movements

References I

Malcolm Slaney.
The 13 db miracle.
https://ccrma.stanford.edu/~malcolm/13dB_Miracle/, 2024.

Kyle Forinash and Wolfgang Christian.
Just noticeable difference.
https://phys.libretexts.org/Bookshelves/Waves_and_Acoustics/Book%3A_Sound_-_An_Interactive_eBook_(Forinash_and_Christian)/07%3A_Pitch_Loudness_and_Timbre/7.01%3A_Pitch_Loudness_and_Timbre/7.1.04%3A_Just_Noticeable_Difference, 2024.

Weiping Tu, Ruimin Hu, Heng Wang, and Wenqin Chen.
Measurement and analysis of just noticeable difference of interaural level difference cue.
In *2010 International Conference on Multimedia Technology*, pages 1–3, 2010.

Taewoong Lee, Jesper Kjær Nielsen, and Mads Græsbøll Christensen.
Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters.
IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28:2412–2426, 2020.

Steven Par, Armin Kohlrausch, Heusdens Richard, Jesper Jensen, and Søren Jensen.
A perceptual model for sinusoidal audio coding based on spectral integration.
EURASIP Journal on Advances in Signal Processing, 2005, 06 2005.

References II

Stefania Cecchi, Laura Romoli, Paolo Peretti, and Francesco Piazza.
Low-complexity implementation of a real-time decorrelation algorithm for stereophonic acoustic echo cancellation.
Signal Processing, 92(11):2668–2675, 2012.

Stefania Cecchi, Alberto Carini, Francesco Piazza, and laura romoli.
A multichannel and multiple position adaptive room response equalizer in warped domain.
09 2013.

Yue Qiao and Edgar Choueiri.
Sann-psz: Spatially adaptive neural network for head-tracked personal sound zones, 2024.