ICASSP Tutorial T1: Personalising sound over loudspeakers

Dr Philip Jackson, CVSSP, U. Surrey, UK
Dr Filippo Fazi, ISVR, U. Southampton, UK
Dr Phil Coleman, IoSR/CVSSP, U. Surrey, UK
Acknowledgements

The following people have made important contributions to the work presented in this tutorial:

Khan Baykaner, Søren Bech, Jon Francombe, Craig Henry, Russell Mason, Martin Møller, Marek Olik, Martin Olsen, Qiaoxi Zhu, Ferdinando Olivieri, Mincheol Shin, Falk-Martin Hoffmann, Eric Hamdan
Introduction to personalising sound over loudspeakers
Why sound?

• In evolutionary terms, **sensing** of sound and vibration pre-dates vision

• Hearing provides **connection** to the world around us, through the grapevine, the jungle drum, the market hubbub, the buzz

• It is especially adept at detecting transient events, as with unexpected sounds that turn the head and grab **attention**

• As a truly panoramic sensing modality, we can interpret **immersive** environments

Moore, 2004. An Introduction to the Psychology of Hearing
Why personal?

- Increasingly common to multi-task during *shared* experiences
- Even in company, *personal* life continues (alerts, notifications)
- We like *choice* to control our experiences, including selective enhancement, content filters or supplementary feeds
- Personalisation improves autonomy, intimacy, and sense of privacy

**Object-based audio** (e.g., MPEG-H) offers many personalisation features

Bleidt et al., 2014. Object-based audio: Opportunities for improved listening..., SMPTE
ITU, 2015. ITU-R BS.2076-0: Audio Definition Model
Herre et al., 2015. MPEG-H 3D Audio - The new standard for coding of immersive spatial audio, JSTSP
Coleman et al., 2018. An audio-visual system for object-based audio: from recording to listening, TMM
Why loudspeakers?

• Loudspeakers can be incorporated into the built environment
  • Clean, safe, convenient

• Absence of wearables (c.f. headphones, earphones, headsets)
  • Not weighed down, good comfort and hygiene, free to move, open to alerts

• Creates space for shared experiences
  • Being together in the space, maintains open communication channel

• Provide consistent spatial and timbral quality
What are the practical challenges?

• Sound is invisible!
• Sources spread sound out in all directions
• Air moves and changes temperature, pressure and humidity
• Environments reflect sound all over the place, can be noisy and vary
• Sound diffracts around obstacles and permeates barriers
• Transducers have noise, colouration, non-linear distortion and frequency/power limits
• Listeners are complex, whose perception changes with context

Kinsler et al., 2000. Fundamentals of Acoustics
Moore, 2004. An Introduction to the Psychology of Hearing
Vision for future ears-free personal sound

The promise of sound is in its power:
• To tell a story
• To transport you to another world
• To envelope you in it
• To transcend the mundane, to evoke emotion, passion and encounter empathy
• To connect with your loved ones and your tribe! grab our attention

To fulfil this, we need capability:
• To deliver you sound that is reliable, enveloping and tailored to your needs
What will we cover in this tutorial?

- **Formulation**
  - Definitions of personal sound zones and spatial audio
  - Key approaches to manipulate sound fields
  - Measures of performance

- **Engineering**
  - Design parameters
  - Regularization
  - Room effects
  - Practical sound zone filter design

- **Experience**
  - Soundbar listening demonstration
  - Perceptual models
  - Alternative approaches and applications

- **Summary of conclusions and perspectives**
Definitions:
Personal sound zone and spatial audio problems
Zone size

**Theatre** *(auditorium)*

**Lounge** *(head)*

**Person** *(ear)*

André et al., 2014, IJHCS

Olik et al., 2013, AES conf

Hollebon et al., 2019, AES conf
Zone count

With $Z$ zones, each bright zone has $Z-1$ dark zones:

<table>
<thead>
<tr>
<th>Bright</th>
<th>Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td></td>
</tr>
<tr>
<td>{A}</td>
<td>{B}</td>
</tr>
<tr>
<td>{A}</td>
<td>{BC}</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>{A}</td>
<td>{BCDEF}</td>
</tr>
</tbody>
</table>
Spatial sound formats

- 1D (distance): mono
- 2D (distance+azimuth): stereo, 5.1 surround
- 3D (distance+azimuth+height): 9.1, 22.2, etc.

- Object-based audio

- Real vs. virtual channels

Rumsey, 2001. Spatial audio
Coleman et al., 2014. Stereophonic personal audio reproduction using planarity control optimization, ICSV
Coleman & Jackson, 2016. Planarity-based sound field optimization for multi-listener spatial audio, AES
ITU, 2015. ITU-R BS.2076-0: Audio Definition Model
Thresh & Kearney, 2017. A direct comparison of localisation performance when using first..., AES
Reproduction system

• Number of channels, \( L \)
• Loudspeaker arrangement:
  • Uniform line array
  • Uniform circular array
  • Arbitrary positions
• Assuming:
  • Calibrated gain
  • Flat full-range frequency response
  • Perfect synchronisation
• Ideal (plane or) monopole source, \( q \)

Møller & Olsen, 2011. Sound zones, MSc thesis
Coleman, 2014. Loudspeaker array processing for personal sound zone reproduction. PhD thesis
Zone sampling

Nyquist principle imposes maximum spacing at half the wavelength of the highest frequency:

\[ d_{\text{max}} = \frac{c}{2f} \]

e.g.,

\[ d_{\text{max}} = 5 \text{ cm at } 3.4\text{kHz} \]

Williams, 1999, Fourier Acoustics
Essential notation

**Sources**

\[ \mathbf{q}(f) = [q_1, \ldots, q_l, \ldots, q_L]^T \]

**Control mics**

\[ \mathbf{p}_A(f) = [p_{A,1}, \ldots, p_{A,n}, \ldots, p_{A,N}]^T \]

**Transfer functions**

\[
\mathbf{G}_A(f) = \begin{bmatrix} G_{A,1,1} & \cdots & G_{A,1,L} \\ \vdots & \ddots & \vdots \\ G_{A,N,1} & \cdots & G_{A,N,L} \end{bmatrix}; \quad \mathbf{G}_B(f) = \begin{bmatrix} G_{B,1,1} & \cdots & G_{B,1,L} \\ \vdots & \ddots & \vdots \\ G_{B,N,1} & \cdots & G_{B,N,L} \end{bmatrix}
\]

**Relations**

\[ \mathbf{p}_A = \mathbf{G}_A \mathbf{q} \quad \text{and} \quad \mathbf{p}_B = \mathbf{G}_B \mathbf{q} \quad \text{or} \quad [\mathbf{p}_A] = [\mathbf{G}_A] [\mathbf{q}] \]

Coleman, 2014. Loudspeaker array processing for personal sound zone reproduction. PhD thesis
Definitions summary

• Number, size and spacing of zones
• Spatial audio objectives
• Reproduction setup
• Control and monitor microphones
• For monopole sources:

\[ G_{n,l} = \frac{j \rho f}{2R} e^{j2\pi fR/c}, \quad \text{where} \ R = |r_{n,l}| \]
Key approaches to filter design for personal sound zones
Notation

- \( q_\ell(\omega) \) driving signal of the \( \ell \)-th loudspeaker
- \( p_m(\omega) \) signal of the \( \ell \)-th microphone/control point
- \( G_{m\ell}(\omega) \) electroacoustical transfer function between the \( \ell \)-th speaker and the \( m \)-th control point

\[
p_m(\omega) = \sum_{\ell=1}^{L} G_{m\ell}(\omega) q_\ell(\omega)
\]
Notation

\[ \mathbf{q} = \left[ q_1(\omega), q_2(\omega), \ldots q_L(\omega) \right]^T \]

\[ \mathbf{p} = \left[ p_1(\omega), p_2(\omega), \ldots p_N(\omega) \right]^T \]

\[ \mathbf{G} = \begin{bmatrix} G_{1,1}(\omega) & \cdots & G_{1,L}(\omega) \\ \vdots & \ddots & \vdots \\ G_{N,1}(\omega) & \cdots & G_{N,L}(\omega) \end{bmatrix} \]

\[ \mathbf{p} = \mathbf{G} \mathbf{q} \]
Interior problem

Exterior problem

Zone 1

Zone 2

Zone 1

Zone 2

Zone 3

Array
Pressure matching

$p_T$ is the target pressure vector in the bright zone

$$p_T = \begin{bmatrix} p_1, p_2, \ldots, p_{N_B}, 0, 0, \ldots, 0 \end{bmatrix}^T$$

$N > L \rightarrow$ overdetermined problem

$$J = \|p - p_T\|^2 = \|G q - p_T\|^2$$

$$q_{opt} = G^\dagger p_T = \left( G^H G \right)^{-1} G^H p_T$$

Relation to acoustical holography

ACOUSTICAL HOLOGRAPHY

Measurement of sound field

The source strength is determined by solving an inverse problem

Vibrating structure that generates and acoustic field

Desired sound field

SOUND FIELD CONTROL

The loudspeaker signals are calculated by solving an inverse problem

Array of loudspeakers

• The (physical) solution exists

• The solution might not exist!
Ill-conditioning

\[ q_{opt} = G^\dagger p_T \quad J = \|G q - p_T\|^2 \]

Matrix \( G \) can be ill-conditioned, especially at low frequencies (columns of \( G \) are almost identical).

Array effort
\[ E(\omega) \propto \|q\|^2 = \sum_{\ell=1}^{L} |q_\ell(\omega)|^2 \]

Effort may be very large if \( G \) is ill-conditioned, leading to unstable solutions.
Tikhonov Regularization

Cost function with Tikhonov regularization

\[ J = \| G \, q - p_T \|^2 + \beta \| q \|^2 \]

\[ q_{opt} = (G^H G + \beta I)^{-1} G^H p_T \]

Reduces of array effort, but increases error

Example: multi-zone with cylindrical array
Reproduced radiation pattern

Radiation pattern at 94 Hz

Radiation pattern at 2016 Hz

Radiation pattern at 9984 Hz

Low freq. limit

High freq. limit

On-axis reproduced pressure
Bright and dark zones

\[ \mathbf{q} = [q_1(\omega), q_2(\omega), \ldots q_L(\omega)]^T \]

\[ \mathbf{p}_B = \left[ p^{(B)}_1(\omega), p^{(B)}_2(\omega), \ldots p^{(B)}_{N_B}(\omega) \right]^T \]

\[ \mathbf{p}_D = \left[ p^{(D)}_1(\omega), p^{(D)}_2(\omega), \ldots p^{(D)}_{N_D}(\omega) \right]^T \]

\[ \mathbf{p}_B = \mathbf{G}_B \mathbf{q} \]

\[ \mathbf{p}_D = \mathbf{G}_D \mathbf{q} \]

\[ \mathbf{p} = \begin{bmatrix} \mathbf{p}_B \\ \mathbf{p}_D \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_B \\ \mathbf{G}_D \end{bmatrix} \]
Acoustic contrast

It is the ratio of the average acoustic potential energy in two zones

\[ C = \frac{< E_B >}{< E_D >} \approx \frac{\| p_B \|^2 / N_B}{\| p_D \|^2 / N_D} \]

\[ = \frac{N_D}{N_B} \frac{q^H G_B^H G_B q}{q^H G_D^H G_D q} \]

\[ p_{B/D} = G_{B/D} q \]
Acoustic contrast maximisation

• Direct formulation: maximises the energy in the bright zone while keeping the energy in the dark zone to a constant value $D$ and the effort below a given value $E$

$$\text{Maximise } \|p_B\|^2 \text{ s.t. } \|p_D\|^2 = D \text{ and } \|q\|^2 \leq E$$

• Indirect formulation: minimise the energy in the dark zone while keeping the energy in the bright zone to a constant value $B$ and the effort below a given value $E$

$$\text{Minimise } \|p_D\|^2 \text{ s.t. } \|p_B\|^2 = B \text{ and } \|q\|^2 \leq E$$

Elliott, S.J., et al., 2012. “Robustness and regularization of personal audio systems”. TASLP.
Acoustic contrast maximisation (direct)

- Maximise $\|p_B\|^2$ s.t. $\|p_D\|^2 = D$ and $\|q\|^2 \leq E$

$L = q^H G_B^H G_B q - \lambda_1 (q^H G_D^H G_D q - D) - \lambda_2 (q^H q - E)$

$
\lambda_1 q_{opt} = \left( G_D^H G_D \right)^{-1} \left( G_B^H G_B - \lambda_2 I \right) q_{opt}
$

- The optimal solution $q_{opt}$ is the eigenvector associated with the largest eigenvalue $\lambda_1$

Elliott, S.J., et al., 2012. “Robustness and regularization of personal audio systems”. TASLP.

Acoustic contrast maximisation (indirect)

• Minimise \( ||p_D||^2 \) s.t. \( ||p_B||^2 = B \) and \( ||q||^2 \leq E \)

\[
L = q^H G_D^H G_D q + \lambda_1 (q^H G_B^H G_B q - B) \\
+ \lambda_2 (q^H q - E)
\]

\[
\lambda q_{opt} = \left( G_D^H G_D + \lambda_2 I \right)^{-1} \left( G_B^H G_B \right) q_{opt}
\]

• The optimal solution \( q_{opt} \) is the eigenvector associated with the largest eigenvalue \( \lambda \)

Elliott, S.J., et al., 2012. “Robustness and regularization of personal audio systems”. TASLP.
Energy difference maximisation

- Maximise energy difference
  \[ \|p_B\|^2 - \alpha \|p_D\|^2 \text{ s.t. } \|q\|^2 \leq E \]

\[ L = q^H G_B^H G_B q - \alpha q^H G_D^H G_D q - \lambda_2 (q^H q - E) \]

\[ \lambda_2 q_{opt} = \left( G_B^H G_B - \alpha G_D^H G_D \right) q_{opt} \]

- The optimal solution \( q_{opt} \) is the eigenvector associated with the largest eigenvalue \( \lambda_2 \).

Shin et al., 2010. “Maximization of acoustic energy difference between two spaces”. JASA
Performance measures
Measures of personal sound quality

• Interference
  • From sounds other than the intended personal sound

• Robustness
  • For any practical implementation

• Timbral quality
  • Quality of the received personal sound

• Spatial quality
  • Reproduction artefacts or spatial distortion

Emiya et al., 2011. Subjective and objective quality assessment of audio source separation, TASLP
Francombe et al., 2014. Elicitation of attributes for the evaluation of audio-on-audio interference, JASA
George et al., 2010. Development and validation of an unintrusive model for predicting..., JAES
Jackson et al., 2008. QESTRAL (Part 3): System and metrics for spatial quality prediction, AES
Goodness of fit: MMSE

- Target sound field
- Given a target sound field, 
  \( \mathbf{d}_A(f) = [d_{A,1}, ..., d_{A,m}, ..., d_{A,M}]^T \),
  the mean squared error is
  \[
  \text{MSE} = \frac{1}{M} \sum_{m=1}^{M} |p_{A,m} - d_{A,m}|^2
  \]
  or
  \[
  \text{MSE} = \frac{1}{M} (\mathbf{p}_A - \mathbf{d}_A)^H (\mathbf{p}_A - \mathbf{d}_A)
  \]

Independent sampling points

Sources \( q(f) = [q_1, ..., q_l, ..., q_L]^T \)

Control mics \( p_A(f) = [p_{A,1}, ..., p_{A,n}, ..., p_{A,N}]^T \)

Monitor mics \( o_A(f) = [o_{A,1}, ..., o_{A,m}, ..., o_{A,M}]^T \)

Transfer functions

\[
G_A(f) = \begin{bmatrix}
G_{A,1,1} & \cdots & G_{A,1,L} \\
\vdots & \ddots & \vdots \\
G_{A,N,1} & \cdots & G_{A,N,L}
\end{bmatrix}; \quad \Omega_A(f) = \begin{bmatrix}
\Omega_{A,1,1} & \cdots & \Omega_{A,1,L} \\
\vdots & \ddots & \vdots \\
\Omega_{A,M,1} & \cdots & \Omega_{A,M,L}
\end{bmatrix}
\]

Relations

\[
[p_A] = [G_A] [q] \quad [o_A] = [\Omega_A] [q]
\]

Coleman, 2014. Loudspeaker array processing for personal sound zone reproduction. PhD thesis
Contrast

- Contrast reflects the ratio of the sound pressure levels (SPLs) in the zones

\[ C = 10 \log_{10} \frac{N_B p_A^H p_A}{N_A p_B^H p_B} \]

\[ C = 10 \log_{10} \frac{M_B o_A^H o_A}{M_A o_B^H o_B} \]

Choi & Kim, 2002. Generation of an acoustically bright zone with an illuminated region..., JASA
Coleman et al., 2014. Acoustic contrast, planarity and robustness of sound zone methods..., JASA
Calibration (control) and evaluation (monitor) microphones.

Jackson & Ross, 1996. Application of active noise control to corporate aircraft, ASME
Møller & Olsen, 2016. Sound zones: On performance prediction..., AES conf
Effort

• Efficiency
  • Total acoustical power
• Robustness
  • Errors in calibration
  • Influence of reflections
• Effort metric:

\[ E = 10 \log_{10} \frac{q^H q}{|q_r|^2} \]

Elliott et al., 2012. Robustness and regularization of personal audio systems, TASLP
Coleman et al., 2014. Acoustic contrast, planarity and robustness of sound zone methods..., JASA
Frequency response

- Timbral quality

- Effective bandwidth

- Flatness across frequency
  - Tone, colouration
  - Comb filtering

PEAQ
Baykaner et al., 2015. The relationship between target quality and interference in sound zone, JAES
Spatial measures

- Estimated angular spectrum

Williams, 1999, Fourier Acoustics
Coleman et al., 2014. Acoustic contrast, planarity and robustness of sound zone methods..., JASA
Planarity

• Summary measure of sound field homogeneity, i.e., vs. plane wave

\[ \eta = \frac{\sum_i w_i u_i \cdot u_\alpha}{\sum_i w_i} \]

where unit vector \( u_i \) points in the \( i \)th look direction and has weight \( w_i \), and principal unit vector \( u_\alpha \) is defined as \( \alpha = \text{argmax}_i w_i \).

Merimaa, 2007. Energetic sound field analysis of loudspeaker reproduction, AES
Jackson et al., 2013. Sound field planarity characterized by superdirective beamforming, POMA
Metrics summary

• Goodness of fit: **Mean squared error**, MSE
  for least-squares optimization formulation

• Interference: **Contrast**, C
  for assessing relative SPL or loudness

• Robustness: **Effort**, E
  as a strong indicator for practical performance

• Bandwidth: **Frequency range**, F
  for application-driven content coverage

• Directivity: **Planarity**, \( \eta \)
  for spatial coherence at the listener
Fundamental Relationships under Ideal Conditions: Cost Function and System Setup
Aims

• Demonstrate the characteristics of ACC and PM:
  • Sound fields reproduced
  • Frequency range of operation

• Demonstrate the influence of loudspeaker array design:
  • Circular array (fully enclosed zones, relatively sparse loudspeaker spacing)
  • Line array (compact, relatively close loudspeaker spacing)
Simulation Setup

- **Dark Zone**: Radius $r_r = 0.9$ m, Centered at a distance $r_c = 1.68$ m from the origin.
- **Bright Zone**: Located at $r = 1$ m from the origin.

Key Angles:
- $\psi = 90$ degrees
- $\psi = 180$ degrees

Circumference PM DOA: 211 degrees

Line: PM DOA from array center

Simulation Setup

- Two arrays: line and circle
  - Circle: 60 loudspeakers, 1.68 m radius
  - Line: 8 loudspeakers, 10 cm spacing
- Two zones
- Regularization: Max matrix condition number of $10^{10}$, Array Effort limit of 0 dB
- Listening level 76 dB SPL
- Acoustic contrast capped at 76 dB (i.e. no audible sound in dark zone)
Acoustic Contrast Control: Circular Array

- Performance over frequency

![Graphs showing performance over frequency for Acoustic Contrast, Planarity, and Effort.](image-url)
Acoustic Contrast Control: Circular Array

- Performance over frequency
Acoustic Contrast Control: Circular Array

• Performance over frequency

![Graphs showing acoustic contrast, planarity, and effort over frequency.](image)
Acoustic Contrast Control: Circular Array

- Sound field: sound pressure level

![Images showing sound pressure level for 500 Hz, 1000 Hz, and 4000 Hz frequencies]
Acoustic Contrast Control: Circular Array

- Sound field: sound pressure level

![Diagram showing sound pressure level at 500 Hz, 1000 Hz, and 4000 Hz]
Acoustic Contrast Control: Circular Array

- Sound field: real part of the sound pressure
Pressure Matching: Circular Array

- Performance over frequency
Pressure Matching: Circular Array

- Sound field: sound pressure level
Pressure Matching: Circular Array

- Sound field: real part of the sound pressure
Performance Comparison: Circular Array

- Control of energy vs control of phase
Summary: Circular Array

- **ACC** controls energy only; **PM** controls energy and phase
  - **ACC** gives excellent contrast
  - **ACC** controls contrast over a wide frequency range
  - **PM** gives excellent control of sound in the bright zone (homogenous, planar)
  - **PM** requires higher effort filters to achieve this control
Acoustic Contrast Control: Line Array

• Performance over frequency

![Graphs showing Acoustic Contrast, Planarity, and Effort over frequency.](image-url)
Acoustic Contrast Control: Line Array

- Sound field: sound pressure level

![Graphs showing sound pressure level at 500 Hz, 1000 Hz, and 4000 Hz](image)
Acoustic Contrast Control: Line Array

- Sound field: real part of the sound pressure
Pressure Matching: Line Array

- Performance over frequency
Pressure Matching: Line Array

- Sound field: sound pressure level

![Graphs of sound pressure level at different frequencies](image-url)
Pressure Matching: Line Array

• Sound field: real part of the sound pressure
Performance Comparison: Line Array

• Control of energy vs control of phase (within array capability)
Summary: Line Array

- **ACC** controls energy only; **PM** controls energy and phase
  - **ACC** gives excellent contrast
    - But only reaches the maximum contrast over a relatively narrow frequency range
  - **ACC** controls contrast over a wide frequency range
    - Extra bandwidth compared to PM not as significant as for circular array
    - Low frequency contrast performance the same for both methods
- **PM** gives excellent control of sound in the bright zone (homogenous, planar)
  - But **ACC** still planar due to limited possible DOAs into zone
- **PM** requires higher effort filters to achieve this control
  - But **ACC** also high effort at most frequencies
Loudspeaker Array Design

- Circular array degradation of acoustic contrast performance
  - **ACC**: onset when aliasing lobe separation equals dark zone diameter

---

![Graphs showing sound pressure level at 5600 Hz, 6400 Hz, and 7200 Hz](image-url)
Loudspeaker Array Design

• Circular array degradation of acoustic contrast performance
  • PM: onset when aliasing lobe begins to impinge on dark zone
Loudspeaker Array Design

• Circular array:
  • ACC moves target beam direction to avoid aliasing for as long as possible
    • Aliasing depends on dark zone size
  • PM target direction specified in optimization
    • Aliasing depends on loudspeaker separation and overall control region size
  • To compare, consider the frequency at which contrast falls below 20 dB

• Recent work by Winter et al. to predict spatial aliasing for SFS

Loudspeaker Array Design

- Line array relationships with wavelength are well known:
  - Array aperture and regularization limit low frequency performance
  - Loudspeaker and zone spacing limit high frequency performance
Loudspeaker Array Design

- So we need a large aperture and close spacing 😐

![Heatmaps for 100 Hz, 1000 Hz, and 6500 Hz showing Sound Pressure Level (dB)]
Summary: Performance Comparison

• **ACC**: creates maximal contrast between the zones, at the cost of having no control over the phase
  • High contrast
  • Low planarity (circular array)
  • Relatively efficient

• **PM**: aims to control the pressure amplitude and phase at each control microphone
  • Lower contrast
  • High planarity
  • Relatively inefficient
  • Weighted versions of PM have also been proposed

• Array geometry imposes fundamental limits on performance
Dealing with room reverberation
Structure of a room impulse response

- **Direct sound**
  - First, clean and narrow
  - From source direction

- **Early reflections**
  - Initially sparse, thickening
  - Absorbed and scattered
  - From all walls, up/down

- **Late reverberation**
  - Dense and diffuse
  - Directionless after mixing time

Kuttruff, 2009. Room Acoustics
Ironing out the earlies

• Early reflections (coherent) colour the response

• Reflections add unwanted copies of loudspeakers as image sources

• For strong low order reflections, it may be beneficial to
  • Reduce radiation in specific directions
  • Maximize effectiveness of room equalization for specific reflection paths

• This can be achieved by geometrical optimization of source positions.


Reduced control effort [Elliott et al., 2012, TASLP] or acoustic power [Jones & Elliott, 2008, JASA]

Array close to zones [Chang et al., 2009, JASA]

Directive sources [Simón Gálvez et al., 2012, JASA]

Superdirective beamforming [Elliott et al., 2010, JASA]
Geometrical considerations

- Arrangement of loudspeakers in relation to reflectors
- An unwanted reflection can be avoided, averted or cancelled
- Higher-order reflections use source modes

Olik et al., 2014. Optimum source geometry for sound zone reproduction with a single reflection, JASA
Late reverberation

There are two main strategies to mitigate effects of late reverberation:

• Truncation during calibration
  By shortening the calibration RIRs used to calculate the personal sound zone filters

• Suppression through regularization
  By applying regularization to make the filters less sensitive to convolutive noise during playback

Olik et al., 2013. A comparative performance study of sound zoning methods in a reflective..., AES
Room effect summary

• Acoustic environments add early reflections and late reverberation

• Left untreated, effects can severely degrade system performance

• Mitigation strategies can be deployed for partial recovery:
  1. Acoustically treat/select the room
  2. Arrange your setup to avoid, avert or cancel reflections
  3. Truncate reverberation from the RIRs in calibration
  4. Apply regularization wisely...
Fundamental Relationships under Non-Ideal Conditions: Robustness and Regularization
Regularization: Fundamentals

• Regularization parameter vs performance (ideal conditions)

• Acoustic Contrast Control:
  • **More regularization = lower effort**
  • Contrast and planarity fairly robust unless over-regularized

Regularization: Fundamentals

• Regularization parameter vs performance (ideal conditions)

• Pressure Matching:
  • More regularization = lower effort
  • Planarity fairly robust unless over-regularized
  • Some local optima in contrast

Regularization: Non-ideal Conditions

- Design challenge: choose optimal regularization parameter
- Some frequencies are more sensitive than others

Zhu, Q, et al., 2017, "Robust Acoustic Contrast Control with Reduced In-situ Measurement by Acoustic Modelling". JAES.
## Regularization: Non-ideal Conditions

- **Design challenge**: choose optimal regularization parameter

<table>
<thead>
<tr>
<th></th>
<th>Approach</th>
<th>Example references</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matrix Perspective</strong></td>
<td>Choose diagonal loading regularization parameter based on condition number / singular values of matrix</td>
<td>Shin, M., et al., 2014. “Controlled sound field with a dual layer loudspeaker array”. <em>J. Sound Vib.</em></td>
</tr>
</tbody>
</table>
Regularization: Non-ideal Conditions

• Design challenge: choose optimal regularization parameter

<table>
<thead>
<tr>
<th>Approach</th>
<th>Example references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix Perspective</td>
<td>Choose diagonal loading regularization parameter based on condition number / singular values of matrix</td>
</tr>
<tr>
<td>Array Effort Perspective</td>
<td>Choose diagonal loading regularization parameter to meet a maximum control effort constraint</td>
</tr>
<tr>
<td></td>
<td>Elliott, S.J., et al., 2012. “Robustness and regularization of personal audio systems”. <em>TASLP.</em></td>
</tr>
<tr>
<td></td>
<td>Coleman, P., et al., 2014. “Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array”. <em>JASA.</em></td>
</tr>
<tr>
<td>Error Modelling Perspective</td>
<td>Choose regularization parameters (matrix) based on predicted system errors</td>
</tr>
<tr>
<td></td>
<td>Zhu, Q, et al., 2017, “Robust reproduction of sound zones with local sound orientation”. <em>JASA-EL.</em></td>
</tr>
<tr>
<td></td>
<td>Zhu, Q, et al., 2017, “Robust Acoustic Contrast Control with Reduced In-situ Measurement by Acoustic Modelling”. <em>JAES.</em></td>
</tr>
</tbody>
</table>
## Regularization: Non-ideal Conditions

- **Design challenge:** choose optimal regularization parameter

<table>
<thead>
<tr>
<th>Approach</th>
<th>Example references</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matrix Perspective</strong></td>
<td>Choose diagonal loading regularization parameter based on condition number / singular values of matrix</td>
</tr>
<tr>
<td><strong>Array Effort Perspective</strong></td>
<td>Choose diagonal loading regularization parameter to meet a maximum control effort constraint</td>
</tr>
</tbody>
</table>
|                                 | Elliott, S.J., et al., 2012. “Robustness and regularization of personal audio systems”. *TASLP.*  
|                                 | Coleman, P., et al., 2014. “Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array”. *JASA.* |
| **Error Modelling Perspective** | Choose regularization parameters (matrix) based on predicted system errors       |
|                                 | Zhu, Q, et al., 2017, ”Robust reproduction of sound zones with local sound orientation”. *JASA-EL.*  
|                                 | Zhu, Q, et al., 2017, ”Robust Acoustic Contrast Control with Reduced In-situ Measurement by Acoustic Modelling”. *JAES.* |
Regularization: Non-ideal Conditions

• Optimal regularization parameter?

<table>
<thead>
<tr>
<th>Approach</th>
<th>Matrix Perspective</th>
<th>Array Effort Perspective</th>
<th>Error Modelling Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choose diagonal loading regularization parameter based on condition number / singular values of matrix</td>
<td>Choose diagonal loading regularization parameter to meet a maximum control effort constraint</td>
<td>Choose regularization parameters (matrix) based on predicted system errors</td>
</tr>
</tbody>
</table>

Zhu, Q, et al., 2017, “Robust Acoustic Contrast Control with Reduced In-situ Measurement by Acoustic Modelling”. JAES.
Regularization: Summary

• Under real-world conditions, regularization is critical:
  • Too little regularization: not robust
  • Too much regularization: no contrast

• Three representative ways to find parameter(s) introduced:
  • Look at the matrix to be inverted
  • Fix the array effort
  • Consider the predicted kind/magnitude of errors in the whole system
Making Sound Zone Filters
Define System Geometry

Set Physical Constraints

Estimate Room Transfer Functions

Calculate Narrowband Source Weights

$q(\omega)$

$q(\omega)$

Parameters

Mono audio

Multichannel audio

$G_A(\omega)$

$G_B(\omega)$

$N_A \times L$

$N_B \times L$

$A, Q$

$L \times 1$
Define System Geometry

Method-specific parameters

Set Physical Constraints

Estimate Room Transfer Functions

Calculate Narrowband Source Weights

Target A

\[ \frac{G_A(\omega)}{N_A \times L} \]

\[ \frac{G_B(\omega)}{N_B \times L} \]

Target B

\[ \frac{q(\omega)}{L \times 1} \]

\[ \frac{q(\omega)}{L \times 1} \]

- Parameters
- Mono audio
- Multichannel audio
Define System Geometry

Estimate Room Transfer Functions

Set Physical Constraints

Method-specific parameters

Calculate Narrowband Source Weights

Target A

Target B

Programme A

Programme B

G_A(\omega)
N_A \times L

G_B(\omega)
N_B \times L

\frac{q(\omega)}{L \times 1}

\frac{L}{1}

\frac{A}{Q}

\frac{L}{1}

\frac{A}{Q}

\frac{L}{1}

Parameters

Mono audio

Multichannel audio
Define System Geometry

Method-specific parameters

Set Physical Constraints

Estimate Room Transfer Functions

Calculate Narrowband Source Weights

Target A

\[ G_A(\omega) \]

\[ G_B(\omega) \]

\[ N_A \times L \]

\[ N_B \times L \]

Target B

\[ q(\omega) \]

\[ L \times 1 \]

Programme A

Programme B

Build FIR Filters

\[ \ast \]

To DAC/loudspeakers

\[ A, Q \]

\[ \ast \]

\[ \ast \]

Parameters

Mono audio

Multichannel audio
Target A
Define System Geometry
Method-specific parameters
Set Physical Constraints
Estimate Room Transfer Functions
Calculate Narrowband Source Weights

Target B

Programme A
Programme B
To DAC/loudspeakers

Demo

\[ q(\omega) \]

\[ G_A(\omega) \]

\[ G_B(\omega) \]

\[ N_A \times L \]

\[ N_B \times L \]

\[ L \times 1 \]

\[ A_Q \]

\[ L \times 1 \]

\[ L \times 1 \]

\[ q(\omega) \]

\[ q(\omega) \]

Parameters
Mono audio
Multichannel audio
Demo: PyZones Toolbox for Sound Zone Filter Design

https://github.com/IoSR-Surrey/PyZones
Demo: multi-zone audio delivery with a 15 channel linear loudspeaker array

With contribution by Eric Hamdan
System geometry
Perceptual Measures of Sound Zone Performance
Does it sound any good then?

• Two main factors contribute:
  • Interference suppression
  • Target sound quality

• Interference from the other zone is affected by:
  • Acoustic contrast (cost function design)
  • Relative loudness / spectra of the programme content (not accounted for in design)

• Sound quality in the target zone is affected by:
  • Flatness of frequency response (constraint, cost function design)
  • Ringing in filters (regularization, cost function design)
  • Bright zone sound propagation (cost function design)
Interference

• Understand the perceptual effects of audio interference
• First need to describe interference in a perceptually relevant way
• Attribute elicitation process led to **distraction** as the most perceptually relevant descriptor

Francombe, J. et al., 2014. “Elicitation of attributes for the evaluation of audio-on-audio interference”. JASA.
Modelling Distraction

• A regression model to predict distraction was developed at Surrey
• Main physical correlates: overall loudness, loudness ratio, interference rejection, and frequency content of the interferer
• Loudness ratio better correlation than (unweighted) target to interferer ratio / acoustic contrast
• Recent work has developed and validated a real-time implementation

Sound Quality

- Frequency response across zone microphones (solid: mean; shaded: min, max)
Sound Quality

• Ringing in filters
• Illustrated with simulated time response at central microphone in zone A
Sound Quality

• Spatial properties of the bright zone

PM: plane wave
ACC: ?
Planarity Control

- PC Cost function
  
  \[
  \text{Min. } J_{PC} = p_B^H p_B + \mu (p_A^H Y_A Y_A p_A - A) + \lambda (q^H q - Q)
  \]
  
  pressure in the dark zone (like ACC)

- Spatial steering matrix
  - Populated by superdirective beamforming

- Specify pass range of directions by means of Gamma term

\[
\Gamma = \text{diag}[\gamma_1, \gamma_2, \ldots, \gamma_I]
\]


Coleman, P., et al., 2014. “Personal audio with a planar bright zone”. *JASA.*
Angular pass range design

• Direction of impinging energy at bright zone
  • **ACC** not constrained: can leads to spatial problems
  • **PM** narrowly constrained
  • **PC** constrained to a range of angles

Coleman, P., et al., 2014. “Personal audio with a planar bright zone”. *JASA.*
Planarity Control

- Performance over frequency

![Graphs showing Acoustic Contrast, Planarity, and Effort across different frequencies and conditions.](image)
Planarity Control

• Sound field: sound pressure level
Planarity Control

- Sound field: real part of the sound pressure
Planarity Control

- Frequency response across zone microphones
  (solid: mean; shaded: min, max)
Perceptual Evaluation

• We ran listening tests to compare ACC, PC and PM
  • Including some variants of PC
• Implemented in the Surrey Sound Sphere (nothing simulated!)
• Listeners separately rated:
  • Target quality
  • Distraction
  • Overall quality

Perceptual Evaluation

• Target Quality

![Graph showing target quality for different sound zoning methods, indicating a trend from worst to best.]
Perceptual Evaluation

- Distraction

![Graph showing distraction levels for different sound zoning methods.](attachment:image)
Perceptual Evaluation

• Overall Quality
Perceptual Evaluation

Conclusions:
- In choosing a cost function, there is a trade-off between target quality and interference.
- The spatial energy distribution and phase in the bright zone was the main difference between methods as tested.
- Planarity control, with a soft DOA constraint, gives a good trade-off between the characteristics of the baseline methods (ACC and PM).
- Other aspects (frequency response, ringing in filters) have not been explored to date in the context of sound zones.
Alternative approaches to personal sound zones
Spherical Harmonic Representation of Sound Fields

A sound field that is a solution to the homogeneous Helmholtz equation can be represented as

\[ p(r, \theta, \varphi, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} B_{nm}(\omega) j_n^m(kr) Y_n^m(\theta, \varphi) \]
The Spherical Harmonics

$n = 1$

$n = 2$

$n = 3$

$m = -3$  $m = -2$  $m = -1$  $m = 0$  $m = 1$  $m = 2$  $m = 3$
Spherical Harmonics and Fourier Series

• Fourier series

• Spherical harmonic expansion (generalised Fourier series)

Target     n=1     n=3     n=5     n=17     n=49
The Spherical Bessel Functions

• The radial dependence of a sound field is expressed by the spherical Bessel functions
Basis functions

\[ i^n j_n(kr) \ Y_n^m(\theta, \varphi) \]
Series truncation

\[ p(r, \theta, \varphi) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} B_{nm} i^n j_n(kr) Y_n^m(\theta, \varphi) \]

\[ N \geq kr \]
Mode matching

Target sound field

\[ p_T(r, \theta, \varphi) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} B_{nm} i^n j_n(kr) Y_n^m(\theta, \varphi) \]

Reproduced sound field

\[ p(r, \theta, \varphi) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \sum_{\ell=1}^{L} H_{nm\ell} q\ell i^n j_n(kr) Y_n^m(\theta, \varphi) \]

Least squares minimisation

\[ J = \| p - p_T \|^2 \quad \Rightarrow \quad J = \sum_{n=0}^{N} \sum_{m=-n}^{n} \sum_{\ell=1}^{L} \left| H_{nm\ell} q\ell - B_{nm} \right|^2 \]

\[ J = \| Hq - b \|^2 \]
Multi-zone mode matching

\[ b_{M \times 1} = \begin{bmatrix} b_A \\ b_B \end{bmatrix}, \quad H_{M \times L} = \begin{bmatrix} H_A \\ H_B \end{bmatrix} \]

\[ M \geq N_z (kr_z + 1)^2 \]

- Least squares minimisation

\[ J = \| H q - b \|^2 \]

\[ q_{opt} = H^+ b = \left( H^H H + \beta I \right)^{-1} H^H b \]

- Conceptually similar to pressure matching

Poletti and Betlehem, 2014. “Creation of a single sound field for multiple listeners”. INTER-NOISE

Zone of silence

Truncated field

\[ p_N = \sum_{n=0}^{N} \sum_{m=-n}^{n} B_{nm} i^n j_n(kr) Y_n^m(\theta, \varphi) \]

Complementary field

\[ p_N^{(C)} = \sum_{n=N+1}^{\infty} \sum_{m=-n}^{n} B_{nm} i^n j_n(kr) Y_n^m(\theta, \varphi) \]
Zone of silence

\[ \mathbf{b}_D = [0, 0, 0, \ldots, B_{M+1,0}, \ldots, B_{NN}]^T \]

\[ M = kr_z, \quad N \geq kr_{\text{max}} \]

- Least squares minimisation

\[ J = \| \mathbf{H} \mathbf{q} - \mathbf{b} \|^2 \]

\[ q_{\text{opt}} = \mathbf{H}^\dagger \mathbf{b} = \left( \mathbf{H}^H \mathbf{H} + \beta \mathbf{I} \right)^{-1} \mathbf{H}^H \mathbf{b} \]

Poletti and Fazi, 2015. “An approach to generating two zones of silence with application to personal sound systems”. JASA
Zone of silence

Poletti and Fazi, 2015. “An approach to generating two zones of silence with application to personal sound systems”. JASA
Cross-Talk Cancellation allows for the delivery of independent signals to the two ears of the listener.
Cross-Talk Cancellation with a loudspeaker array
Multi-listener Cross-Talk Cancellation
Putting it all together: summary & conclusions
Formulation

• Why? For convenient shared experiences

• Design choices
  • Zones, target experience, system setup, calibration & evaluation

• Key approaches:
  • pressure matching (PM), acoustic contrast control (ACC)

• Performance
  • Goodness of fit, contract, effort, bandwidth, planarity
Engineering

• Design-performance axioms
  • Contrast vs. target quality (ACC vs. PM), # sources → bandwidth

• Robust design
  • Frequency-dependent regularization, conditioning/effort/error-model methods, performance-robustness trade off

• Room effects
  • Design & calibrate for the earlies
  • Truncate & regularize for the late

• PyZones interactive sound zone filter design
User experience

• Line-array sound zone listening demonstration
• Perceptual models of performance
  • Sound quality
  • Interference
• Alternative approaches
  • Perceptually-inspired planarity control (PC)
  • Spherical harmonics domain methods
Challenges & opportunities

• Perceptual models
  • Timbral & spatial target quality
  • Distraction & interference
  • Roles of context & of listener goals (attention)

• Adaptive systems
  • Moving listeners
  • Large & dynamic setups
  • Slick & robust calibration and synchronisation

• Paradigm
  • General framework for multi-listener spatial sound
  • Deployment of personalisable content
ICASSP Tutorial: Personalising sound over loudspeakers

Dr Philip Jackson, p.jackson@surrey.ac.uk, CVSSP, U. Surrey, UK
Dr Filippo Fazi, filippo.fazi@soton.ac.uk, ISVR, U. Southampton, UK
Dr Phil Coleman, p.d.coleman@surrey.ac.uk, IoSR/CVSSP, U. Surrey, UK

Thank you for your engagement!

Any questions?
References


Coleman, P., and Jackson, P.J.B., 2014. Planarity panning for listener-centered spatial audio, Audio Engineering Society 55th International Conference on Spatial Audio, Helsinki

Coleman, P., and Jackson, P.J.B., 2016. Planarity-based sound field optimization for multi-listener spatial audio, Audio Engineering Society International Conference on Sound Field Control, Guildford, UK


Hamasaki, K., et al., 2008. A 22.2 Multichannel Sound System for Ultrahigh-Definition TV (UHDTV), SMPTE Motion Imaging Journal
Hollebon, J., Simón Gálvez, M.F., and Fazi, F.M., 2019. Multiple listener crosstalk cancellation using linear loudspeaker arrays for binaural cinematic audio, Audio Engineering Society International Conference on Immersive and Interactive Audio, York, UK
Jackson, P.J.B., Dewhirst, M., Conetta, R., et al., 2008. QESTRAL (Part 3): System and metrics for spatial quality prediction, 125th Audio Engineering Society Conv., San Francisco CA, Preprint 7597


