Virtual Navigation of Ambisonics-Encoded Sound Fields Containing Near-Field Sources

Joseph G. Tylka
Final Public Oral Examination
Mechanical and Aerospace Engineering
Princeton University
Advisor: Edgar Y. Choueiri

9 May 2019
Virtual Navigation
Virtual Navigation
Virtual Navigation
Ambisonics Encoding & Binaural Rendering

Capture ➔ Encode to Ambisonics ➔ Decode to virtual speakers ➔ Convolve ➔ Binaural signals

https://developers.google.com/vr/concepts/spatial-audio

Xie (2013), Head-Related Transfer Function and Virtual Auditory Display, Fig. 2.5.

mh acoustics Eigenmike

https://en.wikipedia.org/wiki/Ambisonics
Near-Field Sources

Sound source

Ambisonics mic.

Region of validity
Virtual Navigation of Ambisonics-Encoded Sound Fields Containing Near-Field Sources

- Several navigational methods exist
  - Unclear how they perform, how to compare
  - When should each method be used?
- Region of validity is a well-known mathematical issue
  - Unclear how it manifests in terms of audio
Outline

A. Metrics for evaluating navigational methods
   • Auditory models and objective metrics
     - Spectral coloration and perceived source localization
   • Subjective validation through listening experiments

B. Effect of microphone validity
   • Proposed method: valid microphone interpolation (VMI)
   • Simple benchmark: weighted-average interpolation (WAI)

C. Comparisons of navigational methods
   • State-of-the-art: time-frequency analysis (TFA)
   • Practical considerations

Summary and conclusions
A. Metrics for Evaluating Navigational Methods
Listening Experiments
Experiment 1: Coloration
Coloration Test

- Collect subjective ratings of coloration & compute objective metrics
- Perform multiple linear regression between ratings and metrics values
- **MULTiple Stimuli with Hidden Reference and Anchor** (ITU-R BS.1534-3)
  - **Reference**: no navigation, pink noise
  - **Anchor**: 3.5 kHz low-passed version of Ref.
- **Test samples**: vary navigational method and distance
- User rates each sample from 0–100: 100 = Ref.; 0 = Anchor
  - Coloration score = 100 – MUSHRA rating: 0 = Ref.; 100 = Anchor
Coloration GUI

Round 1 of 3

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>Good</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>Fair</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>Poor</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td>Bad</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td>Ratings</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>0.0 dB</td>
</tr>
</tbody>
</table>


Save and Proceed
Coloration Metric

- Averaged spectral distortions in narrow frequency bands (Schärer and Lindau, 2009)

- Critical bands approximated by ERB-spaced gammatone filter bank

\[ \eta(f_c) = 10 \log_{10} \left( \frac{\int |H_{\Gamma}(f; f_c)||A(f)|^2 df}{\int |H_{\Gamma}(f; f_c)||A_0(f)|^2 df} \right) \]

\[ \rho_{\eta} = \max_c \eta(f_c) - \min_c \eta(f_c) \]

- More sophisticated coloration metrics exist (Boren et al., 2015)

Regression Results

Test details:
- 27 test samples
- 4 trained listeners
- Pink noise signal

Regression Results

**Proposed:** $r = 0.84$

**Kates:** $r = 0.72$

**Pulkki et al.:** $r = 0.77$

**Wittek et al.:** $r = 0.77$

Experiment 2: Localization
Localization Test

… 10 11 12 13 14 15 …

10 cm

127 cm

5 cm

θ

Recording/encoding

Interpolation
### Localization Test

#### Round 1 of 5

<table>
<thead>
<tr>
<th>Trial</th>
<th>Samples</th>
<th>Head Azimuth</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>2</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>3</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>4</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>5</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>6</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>7</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>8</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>9</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>10</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>11</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>12</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>13</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
<tr>
<td>14</td>
<td>Play</td>
<td>Capture</td>
<td>Az: 0.00 deg</td>
</tr>
</tbody>
</table>

[Image of experimental setup]
Localization Metric

1. Transform to plane-wave impulse responses (IRs)
2. Split each IR into wavelets
3. Threshold to find onset times
4. Compute average amplitude in each critical band
5. Compute perceptually-weighted “energy vector” in each band
6. Compute average vector over all bands

Localization Test Results

Test details:
• 70 test samples
• 4 trained listeners
• Speech signal

Legend
Data/model

\[ y = x \pm 5^\circ \]

Pearson correlation coefficient: \( r = 0.85 \)
Mean absolute error: \( \varepsilon = 3.67^\circ \)

Summary - Part A.

- Developed objective metrics for coloration and localization
- Constructed experimental setup to conduct listening tests
- Conducted subjective listening tests to validate the metrics

- **Finding**: the chosen coloration metric is a dominant factor in human perception of coloration

- **Finding**: the structure of the localization metric is valid for giving reasonable predictions of perceived localization
B. Effect of Microphone Validity
Ambisonics Interpolation

- Sound source
- Amb. mic. 1
- Amb. mic. 2
- Amb. mic. 3
- Listening position
Ambisonics Interpolation

Region of validity

Sound source

Amb. mic. 1

Listening position

Amb. mic. 2

Amb. mic. 3
Ambisonics Interpolation

Region of validity

Sound source

Amb. mic. 1

Listening position

Amb. mic. 3

Amb. mic. 2
Ambisonics Interpolation

- Sound source
- Amb. mic. 1
- Amb. mic. 2
- Amb. mic. 3
- Listening position

Region of validity
Ambisonics Interpolation

Region of validity

Amb. mic. 1

Amb. mic. 2

Amb. mic. 3

Sound source

Listening position
Valid Microphone Interpolation

Amb. signals from mic. 1

Amb. signals from mic. P

Detect and locate near-field sources

Discard signals from invalid microphones

Interpolated amb. signals

Determine valid microphones

Apply interpolation filter matrix

Compute interpolation filter matrix

Microphone positions

Listening position

Numerical Simulations

\[ \gamma = \frac{s_0}{\Delta/2} \]

Results - Coloration

Weighted-Average Interpolation (WAI)  Valid Microphone Interpolation (VMI)

Finding: excluding the invalid microphone improves coloration performance for interior sources

Results - Localization

Weighted-Average Interpolation (WAI)  Valid Microphone Interpolation (VMI)

**Finding**: excluding the invalid microphone improves localization performance for interior sources

Summary - Part B.

• Developed the valid microphone interpolation (VMI) method which excludes invalid microphones from the interpolation

• Compared this method to a simple benchmark: weighted-average interpolation (WAI)

• **Finding**: excluding the invalid microphone significantly improves coloration and localization performance for interior sources
C. Comparisons of Navigational Methods
Time-Frequency Analysis

Source Triangulation

Point-Source Modeling

Thiergart et al. (2013). “Geometry-Based Spatial Sound Acquisition Using Distributed Microphone Arrays.”
Results - Coloration

**Finding:** VMI achieves superior coloration performance for interior sources and/or large array spacings

Results - Localization

**Time-Frequency Analysis (TFA)**

**Valid Microphone Interpolation (VMI)**

**Finding:** TFA achieves superior localization performance only for interior sources with large array spacings

Domains of Practical Application

**Coloration**

Distant Sources

Dense Array

Sparse Array

Intimate Sources

**Localization**

Distant Sources

Dense Array

Sparse Array

Intimate Sources

Summary - Part C.

• Compared the performance of state-of-the-art methods across ranges of practically-relevant conditions

• **Finding**: for applications with intimate sources and sparsely distributed microphones, only VMI yields high tonal fidelity

• **Finding**: in those applications, TFA yields accurate and superior localization

• **Finding**: for applications with distant sources and sparse microphones, VMI and WAI achieve both accurate tonality and localization
Summary

• Developed and experimentally validated objective metrics for evaluating navigational methods

• Developed the *valid microphone interpolation* method which circumvents the region of validity restriction

• Characterized and compared this and other state-of-the-art navigational methods

• Identified practical guidelines for choosing between them based on intended application
Conclusions

I. The proposed metrics give usable predictions of perceived coloration and localization

II. Excluding invalid microphones improves coloration performance and localization accuracy

III. For applications with sparsely distributed microphones:

• For distant sources, both VMI and WAI give high tonal fidelity and accurate localization performance

• For intimate sources, only VMI gives high tonal fidelity and only TFA gives accurate localization performance
Acknowledgements

SONY

FOCAL®
LISTEN BEYOND

mh acoustics

Local 315
Thank You