Evaluation of techniques for navigation of higher-order ambisonics

Acoustics ’17 Boston
Presentation 1pPPb4
June 25th, 2017

Joseph G. Tylka (presenter)
and Edgar Y. Choueiri

3D Audio and Applied Acoustics (3D3A) Laboratory
Princeton University
www.princeton.edu/3D3A
Sound Field Navigation

- HOA mic. 4
- HOA mic. 3
- Sound source
- HOA mic. 2
- HOA microphone
Sound Field Navigation

- Lots of different ways to navigate:
  - Plane-wave translation (Schultz & Spors, 2013)
  - Spherical-harmonic re-expansion (Gumerov & Duraiswami, 2005)
  - Linear interpolation/“crossfading” (Southern et al., 2009)
  - Collaborative blind source separation (Zheng, 2013)
  - Regularized least-squares interpolation (Tylka & Choueiri, 2016)

- Need a way to evaluate and compare them
  - Isolate navigational technique from binaural/ambisonic rendering
  - Subjective testing can be lengthy/costly → **Objective Metrics**
Overview

• For each quality (localization and coloration):
  • Existing metrics
  • Proposed metric
  • Listening test
  • Results

• Summary and outlook
Source Localization
Existing Metrics

• Binaural models:
  • Lindemann (1986); Dietz et al. (2011); etc.
    • Predict perceived source azimuth given binaural impulse responses (IRs)

• Localization vectors:
  • Gerzon (1992) — for analyzing ambisonics
    • Low frequency (velocity) and high frequency (energy) vectors
    • Predict perceived source direction given speaker positions & gains
  • Stitt et al. (2016)
    • Incorporates precedence effect to Gerzon’s energy vector
    • Model requires: direction-of-arrival, time-of-arrival, and amplitude for each source
  • Tylka & Choueiri (2016) generalized algorithm for ambisonics IRs
Proposed 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 Stitt’s energy vector in each band for $f \geq 700$ Hz
6. Similarly, compute velocity vector in each band for $f \leq 700$ Hz
7. Compute average vector weighted by stimulus energies in each band
Localization Test

... 10 11 12 13 14 15 ...

10 cm

5 cm

127 cm

θ

Recording/encoding

Interpolation

10 cm
Localization Test Results

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

Pearson correlation coefficient: $r = 0.77$

Mean absolute error: $\varepsilon = 3.67^\circ$
Spectral Coloration
Existing Metrics

- Auditory band error (Schärer & Lindau, 2009); peak and notch errors (Boren et al., 2015)

- Central spectrum (Kates, 1984; 1985)

- Composite loudness level (Pulkki et al., 1999; Huopaniemi et al., 1999)

- Internal spectrum and $A_0$ measure (Salomons, 1995; Wittek et al., 2007)
Methodology

• Perform multiple linear regression between ratings and various metrics
  • For spectral metrics: compute max–min & standard deviation

• **MUltiple Stimuli with Hidden Reference and Anchor (ITU-R BS.1534-3)**
  • **Reference**: no navigation, pink noise
  • **Anchor 1**: 3.5 kHz low-passed version of **Ref**.
  • **Anchor 2**: +6 dB high-shelf above 7 kHz applied to **Ref**.

• **Test samples**: vary interpolation technique and distance

• User rates each sample from 0–100: 100 = **Ref**.; 0 = **Anchor 1**
  • Coloration score = 100 – MUSHRA rating: 0 = **Ref**.; 100 = **Anchor 1**

• **Proposed model**: auditory band and notch errors only (Boren et al., 2015)
Regression Results

**Proposed: \( r = 0.84 \)**

**Kates: \( r = 0.72 \)**

**Pulkki et al.: \( r = 0.79 \)**

**Wittek et al.: \( r = 0.77 \)**

Legend

- ○ Data/model
- \( y = x \)
- \( y = x \pm 20 \)
Summary and Outlook

• Presented objective metrics that predict localization and coloration

• Validated through comparisons with subjective test results

Next Steps:

1. Compare localization metric with binaural models

2. Validate metrics for other stimuli, directions, conditions

3. Verify generalization to other binaural rendering techniques
References

• Boren et al. (2015). “Coloration metrics for headphone equalization.”
• Huopaniemi et al. (1999). “Objective and Subjective Evaluation of Head-Related Transfer Function Filter Design.”
• Kates (1985). “A central spectrum model for the perception of coloration in filtered Gaussian noise.”
• Lindemann (1986). “Extension of a binaural cross-correlation model by contralateral inhibition.”
• Schultz and Spors (2013). “Data-Based Binaural Synthesis Including Rotational and Translatory Head-Movements.”
• Southern, Wells, and Murphy (2009). “Rendering walk-through auralisations using wave-based acoustical models.”
• Wittek et al. (2007). “On the sound colour properties of wavefield synthesis and stereo.”

Acknowledgments

• Binaural rendering was performed using M. Kronlachner’s ambiX plug-ins: http://www.matthiaskronlachner.com/?p=2015
• The em32 Eigenmike by mh acoustics was used to measure the HOA RIRs: https://mhacoustics.com/products#eigenmike1
• Auditory filters were generated using the LTFAT MATLAB Toolbox: http://ltfat.sourceforge.net/
• P. Stitt’s energy vector code can be found here: https://circlesounds.wordpress.com/matlab-code/