Perceptual Evaluation of Numerical Auditory Scene Synthesis Using Loudspeaker Arrays

Ismael Nawfal, Joshua Atkins, Daniele Giacobello, and Stephen Nimick
Beats Electronics

Introduction

- There are many methods used to achieve a spatial sound field, such as Loudspeaker Binaural Rendering (LBR) (1), Wave-field Synthesis (WFS) (2), Vector-base Amplitude Panning (VBAP) (3), Higher Order Ambisonics (HOA) (4), and Equivalent Source Method (ESM) (5).
- There is limited literature on the perceptual evaluation of spatial sound synthesis methods (6).
- We introduced numerical auditory scene synthesis (NASS) in (7); a flexible numerical method that allows for broadband filter design and the incorporation of perceptual error.
- We present evaluations of timbral and spatial quality using variations of the NASS method for the task of simulating a single source outside the aperture of an 8 speaker array.

1 Methodology

- \(N_1, N_2, N_3\): lengths of the acoustic path, filter, and desired response, respectively.
- \(D, S, M\): modeling delay, number of speakers, and number of target points, respectively.
- \(Z\) and \(W\) represent spatio-temporal transforms.
- \(p, q\) represent the cost function norm, constraint norm, and constraint threshold, respectively.
- \(h = [h_1, h_2, \ldots, h_M]^T\) is the acoustic filter, \(t\) is the target response.
- Filters designed for 8 channel uniform linear array.
- \(G\) and \(t\) are represented by measured HRTF or a spherical wave propagation model.
- The following HRTFs and spherical wave based configurations were evaluated:
  - HRTF, \(q = 2, M = 2\) (HRTF2_L2)
  - HRTF, \(q = \infty, M = 2\) (HRTF2_L2)
  - HRTF, \(q = \infty, M = 12\) (HRTF12_L2)

2 Evaluation

- Underdetermined cases are not spatially robust; the filters are optimized for the center position.
- The HRTF underdetermined cases closely match the expected ear responses at the central position.
- Spherical wave methods, though generating the expected acoustical waveform, do not achieve the desired responses.
- In overdetermined cases, filters are optimized for a larger spatial region resulting in increased error.

Figure X

- Measurement and simulation setup.

2.1 Objective Evaluation

- 13 listeners: 9 experts and 4 naive.
- Five audio excerpts were evaluated: castanets, pink noise, music, male voice, and female voice.

2.2 Subjective Evaluation

- Two tasks:
  - Array and reference speaker in anechoic room.
  - Array and reference speaker in reverberant room.

- Anchor is decorrelated and low-pass filtered. MUSHRA evaluations conducted on headphones.

References