Investigating the Smoothness of Moving Sources Reproduced with Panning Methods

Matthieu Kuntz, Bernhard U. Seeber
Audio Information Processing, Technical University of Munich
E-mail:{matthieu.kuntz, seeber}@tum.de

Introduction

The accurate loudspeaker-based reproduction of a sound field over a given area becomes more difficult as the frequency increases, manifesting itself in higher reproduction errors [1] and poorer sound quality due to stronger comb filtering and erratic interaural cues away from the center of the loudspeaker array [2]. Previous measurements and simulations of reproduction accuracy showed that Higher-Order Ambisonics (HOA) yields the lowest sound pressure level errors up to 1.5-2 kHz, considering a sweet-spot size of about 50 cm [1]. Above 2 kHz nearest loudspeaker mapping (NLS) performed better, followed by Vector Base Amplitude Panning (VBAP, [3]). This led to the recommendation of combining Higher-Order Ambisonics at low frequencies with nearest loudspeaker mapping at high frequencies. Informal listening tests showed that this hybrid panning method is promising for static virtual sources. This study investigates the perceived reproduction quality for moving sources in the free field.

Coloration has already been investigated for synthetic sources under the scope of room acoustics and coloration induced by reflections [4-6] and under the scope of virtual source auralization for static sources [7] over binaural reproduction [8-10]. While the binaural reproduction was shown to not affect coloration differences [8], the use of non-individual HRTFs, their interpolation, and the variance in headphone placement on the head are still likely to affect coloration perception. While the mechanism of binaural decoloration can reduce the perceived coloration in static scenarios, it is unclear how it performs for moving sources [11]. The authors are only aware of one other study investigating the coloration of moving virtual sources using standard broadcasting loudspeaker setups [12].

The present study investigates the amount of dynamic timbral artifacts in moving sources for nearest loudspeaker mapping, Vector Base Amplitude Panning and Perceptually Equalized Panning (PEP, [13]) and their combination with Higher-Order Ambisonics at low frequencies. A pink noise source was auralized in a horizontal arc from -30° to +30. Participants were asked to rate the amount of dynamic timbral artifacts of the sound source for different angular velocities, movement directions and cross-over frequencies.

Methods

The experiment was run in the SOFE loudspeaker array in the anechoic chamber of AIP TUM [14], using the horizontal ring of 36 loudspeakers with a 10° azimuth spacing, equalized in level, delay and phase.

Panning methods

Nearest loudspeaker mapping

The source signal is played by the loudspeaker with the lowest azimuthal distance to the virtual source. Each loudspeaker signal is faded in and out with a 50 ms Hann window.

Vector Base Amplitude Panning

VBAP [3] is an amplitude panning method, the loudspeaker gains depending on the source position between a given loudspeaker pair. They are normalized to yield equal power across all panning directions, which leads to level errors ranging up to 3 dB in the center of the loudspeaker pair when loudspeakers are equalized in phase.

Perceptually Equalized Panning

The Perceptually Equalized Panning method [13] introduces a short filter to correct panning errors occurring with VBAP or Blumlein panning. It uses HRTFs, e.g. simplified ones from a spherical head model, to compute a differential filter between the HRTF at the virtual source position and the HRTFs at the loudspeaker positions used for VBAP playback. This results in a correction of the 3 dB summation error and a timbre equalization at high frequencies. Panned static sources were found to be almost undistinguishable from a reference for 7.5° loudspeaker spacing for broadband noise up to 10 kHz [13].

Higher-Order Ambisonics

17^{th}-order 2D Higher-Order Ambisonics was implemented using the *basic* decoder as described in [15].

Hybrid panning methods

These methods use a combination of HOA to reproduce low frequencies and either NLS, VBAP or PEP to reproduce high frequencies. The source signal is split into two frequency bands using brick wall filtering in the frequency domain. Loudspeaker signals are computed for both bands individually and summed up. They are referred to by adding the prefix 'HOA-' to the panning method used at high frequencies.

Stimuli

The stimulus played from the virtual sound source was pink noise bandlimited between 100 Hz and 15 kHz with a Gaussian fade-in and fade-out time of 50 ms. The virtual source was moving along a circular trajectory between -30° and 30° or vice-versa at constant angular speed. Throughout the experiment, we varied the movement direction (clockwise and counter-clockwise), the cross-over frequency (1.5 kHz, 2 kHz, 4 kHz, 8 kHz) for the hybrid methods and the angular source speed (10°/s, 20°/s, 30°/s, 60°/s, 90°/s).

In addition to the hybrid panning methods, the regular panning methods were also tested for varying angular source speeds and movement directions.

We applied random level roving of ±3 dB around 60 dB SPL to all our stimuli, and each condition was repeated 7 times.

Experimental procedure

Participants were seated in the center of the loudspeaker array and instructed not to move. They were instructed to rate the amount of dynamic timbral artifacts of the sound as the source moves on a 7-point scale ranging from 1 (inaudible) to 7 (extremely audible) using a computer keyboard. The total of 1120 trials were split into 8 runs of 140 trials, lasting around 10-12 minutes each, with a short break between trials. A familiarization session was run before the experiment and after longer breaks to ensure participants were acquainted with the rating scale and the extent of dynamic timbral artifacts they would experience during the experiment. Six self-reported normal hearing subjects took part in this study.

Results

Movement direction

As expected, there was no significant difference in the ratings between the movement directions. Therefore, we averaged the ratings for both directions and present the following results as movement direction independent. Each data point is computed by averaging over the 14 repetitions of each condition.

Differences across reproduction methods

Figure 1 represents the mean ratings, grouped across angular source speeds and cross-over frequencies.

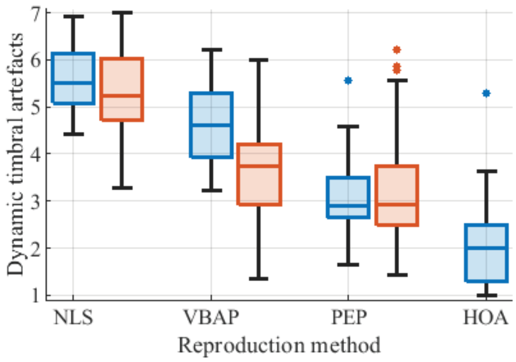

Figure 1: Median and quartiles of the dynamic timbral artefacts ratings for the regular panning methods (blue, n=30) and the hybrid panning methods, combining HOA and a different panning method, indicated on the *x*-axis (red, n=120).

Pure panning methods

As observed in the blue data in figure 1, NLS showed the strongest dynamic timbral artefacts, indicating a very poor auralization of a moving source. VBAP shows a significant improvement, but still yields considerable artefacts, with a median value of 4.6. With a median value of 2.9, we see that PEP performs better than VBAP for the auralization of moving sources. HOA was rated best with a median rating of 2.0. All these differences were deemed significant ($p<0.05$) after multiple pair-wise comparisons using Bonferroni-corrected t-tests.

Hybrid panning methods

The red data in figure 1 show the dynamic timbral artefact ratings for the hybrid panning methods. Similarly to the regular panning methods case, HOA-NLS exhibited the largest artefacts (median of 5.3), followed by HOA-VBAP (median of 3.8), and then HOA-PEP (median of 2.9). The difference between HOA-VBAP and HOA-PEP was also found to be significant ($p<0.05$).

Influence of angular source speed

Pure panning techniques

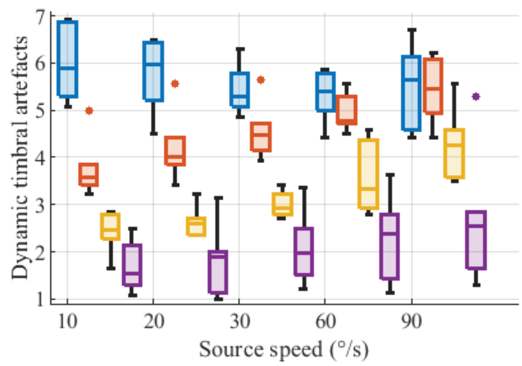

Figure 2: Median and quartiles of the dynamic timbral artefacts ratings for the regular panning methods at different angular source speeds. Blue: NLS; red: VBAP; yellow: PEP; purple: Higher-Order Ambisonics.

Figure 2 represents the dynamic timbral artefact ratings for different angular source speeds. NLS shows a slight tendency towards lower artefact ratings at higher angular speeds, as opposed to the other panning methods. The effect of angular speed on HOA is detrimental, but also quite limited. VBAP and PEP on the other hand show strong deterioration as the angular source speed increases, their median ratings increasing respectively from 3.6 to 5.5 and from 2.5 to 4.3.

Hybrid panning techniques

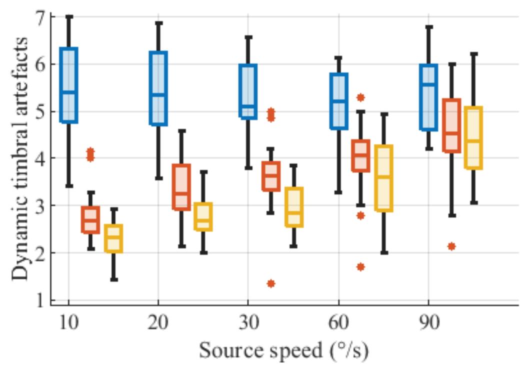

Figure 3: Median and quartiles of the dynamic timbral artefacts ratings for the hybrid panning methods at different angular source speeds. Blue: HOA-NLS; red: HOA-VBAP; yellow: HOA-PEP.

The effect of angular source speed on the hybrid panning methods is similar to that of the regular panning methods, barely affecting HOA-NLS and strongly increasing perceived artefacts for HOA-VBAP and HOA-PEP, with median ratings going respectively from 2.7 to 4.5 and from 2.3 to 4.4.

Influence of cross-over frequency in hybrid panning methods

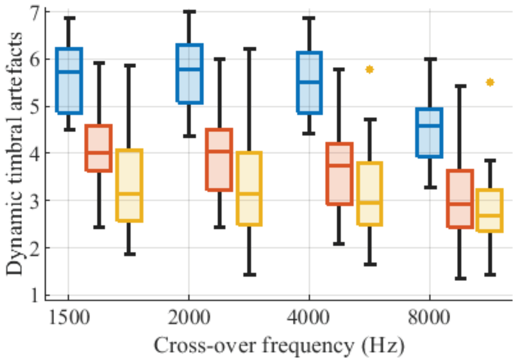

Figure 4: Median and quartiles of the dynamic timbral artefacts ratings for the hybrid panning methods for different cross-over frequencies. Blue: HOA-NLS; red: HOA-VBAP; yellow: HOA-PEP.

As shown in figure 4, the influence of the cross-over frequency on timbral artefacts is limited. For all hybrid panning methods, the ratings were not found to be significantly different across panning methods except for a cross-over frequency of 8 kHz, which showed a significant difference to all other conditions.

Discussion

The results showed timbral artefact ratings of moving virtual sources, auralized with different panning techniques and combinations thereof.

The ratings of the regular panning methods matched our expectations: NLS exhibits the highest artefacts, followed by VBAP, then PEP and then HOA. We verified that the PEP filters significantly improve VBAP, but the 2-loudspeaker panning still contains some timbral changes along the sources' trajectory.

Increasing angular source speeds affected the perceived artefacts most with VBAP and PEP. These methods introduce a change in timbre and source width when the panning direction changes from the same direction as a loudspeaker (effectively equating to nearest loudspeaker mapping) and between a loudspeaker pair. With increasing speed, the rate of these variations increases from 1 Hz (10°/s) to 9 Hz (90°/s), which negatively impacts the artefact perception along the trajectory. The slight reduction in perceived artefacts with NLS and increasing speeds is due to the short fade in and fade out times of the loudspeaker signals, which tends towards amplitude panning methods at high speeds.

While sound pressure measurements show an ideal cross-over frequency of 1.5 kHz, perceived artifacts in moving sources are almost unaffected by the cross-over frequency. This is a strong indication that the ratings were driven by high frequency artefacts.

The similar ratings of HOA-PEP and PEP suggest that PEP and HOA are equally suited to reproduce moving sources in the center of the loudspeaker array.

Conclusion

This study investigates the amount of dynamic timbral artefacts in moving sources auralized by different panning methods. Additionally, combinations of Higher-Order Ambisonics and either nearest loudspeaker mapping, Vector Base Amplitude Panning and Perceptually Equalized Panning were tested. We found that the use of Perceptually Equalized Panning provides a decrease in perceived artefacts compared to Vector Base Amplitude Panning. We found that higher angular source speeds lead to stronger artefacts, as the rate of changes between on-loudspeaker and between-loudspeaker panning increased. Combining Higher-Order Ambisonics with another panning technique at high frequencies does not lead to lower artefact ratings. In fact, they were lowest for HOA.

It is important to note that this study was run for static participants, seated in the center of the loudspeaker array, where Higher-Order Ambisonics is most accurate. As participants step away from the center, Higher-Order Ambisonics quickly exhibits strong comb filter structures, introducing high artefacts in the sound field. This is also worsened when combining the off-center positioning with head movements, especially translations. Under these conditions the panning methods Vector Base Amplitude Panning and Perceptually Equalized Panning are more robust and less coloured. Future work will look into the perception of artefacts in the auralization of moving sound sources at non-optimal listening positions in a loudspeaker array.

Acknowledgments

The SOFE system was financed by the BMBF Bernstein Center for Computational Neuroscience, 01GQ1004B. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 352015383 – SFB 1330 C 5.

Literature

[1] Kuntz, M., Kolotzek, N., Seeber, B.U.: Gemessener Schalldruckpegel im Lautsprecherarray für verschiedene Schallfeldsyntheseverfahren. Fortschritte der Akustik (2021), 1437-1440

[2] Kuntz, M., Seeber, B.U.: Sound field Synthesis: Simulation and evaluation of interaural cues over an extended area. Proceedings of the Euronoise 2021 Conference (2021), 1830-1839

[3] Pulkki, V.: Virtual Sound Source Positioning Using Vector Base Amplitude Panning. Journal of the Audio Engineering Society 45(6) (1997), 456-466

[4] Salomons, A.M.; Coloration and binaural decoloration of sound due to reflections. PhD Thesis, Technical University Delft (1995)

[5] Buchholz, J.: A quantitative analysis of spectral mechanisms involved in auditory detection of coloration by a single wall reflection. Hearing Research 277(1-2) (2011), 192-203

[6] Boussard, P., Santon, F.: Perception De La Coloration D'un Bruit Blanc Due À Un Écho Latéral. Le Journal de Physique IV 02(C1) (1992), C1-221-C1-224

[7] Pulkki, V.: Coloration of Amplitude Panned Virtual Sources. 110th AES Convention (2001), 5402

[8] Wierstorf, H.: Perceptual assessment of sound field synthesis. PhD Thesis, Fakultät IV – Elektrotechnik und Informatik, Technische Universität Berlin (2014)

[9] Wittek, H.: Perceptual differences between wavefield synthesis and stereophony, PhD Thesis, Department of Music and Sound Recording School of Arts, Communication and Humanities, University of Surrey (2007)

[10] Winter, F., Wierstorf, H., Hold, C., Kruger, F., Raake, A., Spors, S.: Colouration in Local Wave Field Synthesis IEE/ACM Transactions on Audio, Speech, and Language Processing 26(10) (2018), 1913-1924

[11] Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., Zotter, F.: Spatial Sound With Loudspeakers and Ist Perception: A Review of the Current State. Proceedings of the IEEE 101(9) (2013), 1920-1938

[12] Romanov, M., Zotter, F., Frank, M., Pike, C.: Comparison of spatial audio rendering methods for ITU-R BS. 2051 standard loudspeaker layouts. International Conference on Spatial Audio (2017)

[13] Seeber, B.U., Hafter, E.: Perceptual equalization of artifacts of sound reproduction via multiple loudspeakers. Meetings on Acoustics (2013), 050045

[14] Seeber, B.U., Clapp, S.: Interactive simulation and free-field auralization of acoustic space with the rtSOFE. The Journal of the Acoustical Society of America 141(5) (2017), 3974

[15] Zotter, F., Frank, M.: Ambisonics: A Practical 3D Audio Theory for Recording. Springer, 2019.