Real-time binaural rendering with virtual vector base amplitude panning

Rishi Shukla, Iuliu Teodor Radu, Mark Sandler, and Rebecca Stewart

Centre for Digital Music, Queen Mary University of London, London, UK

Correspondence should be addressed to This paper was presented at the Audio Engineering Society Conference on Immersive and Interactive Audio, as paper number PS-25. The full published version can be found at: (http://www.aes.org/e-lib/browse.cfm?elib=20414)

ABSTRACT

We present a virtual vector base amplitude panning (VBAP) implementation for 3D head-tracked binaural rendering on an embedded Linux system. Three degrees of freedom head-tracking is implemented within acceptable levels of latency and at 1° angular resolution. The technical performance of virtual VBAP is evaluated alongside a First Order Ambisonics (FOA) approach on the same platform, using analysis of localisation cue error against a human-measured head-related transfer function set. Our findings illustrate that, in scenarios utilising embedded or other portable, low-resource computing platforms, the nature and requirements of the immersive or interactive audio application at hand may determine whether virtual VBAP is a viable (or even preferable) approach compared to virtual FOA.

1 Introduction

In recent years, spatial audio formats have attracted increased commercial interest as a means of delivering immersive media content. The growth of virtual, augmented and mixed reality systems and 360 degree video has encouraged focus on Ambisonics as a means of rendering 2D and 3D sound. For example, Google and Facebook both provide Ambisonic encoding tools for media production [1, 2], whilst game development platform Unity enables integration with Ambisonic decoding plugins [3]. Historically, B-Format Ambisonics encoding was developed in parallel with associated recording techniques and microphone technologies aimed at transparent and faithful capture of real world sound fields [4, 5]. Ambisonics therefore lends itself well to the aim of (re)creating immersive environments founded on representation of true acoustic spaces and is understandably favoured for alternate reality, gaming and cinematic applications.

Virtual Ambisonics describes the technique of rendering 2D or 3D B-Format audio over headphones by synthesising the position of virtual loudspeakers binaurally using HRTF (head-related transfer function) sets. HRTFs usually manifest as multiple left/right pairs of head related impulse responses (HRIRs) measured at desired spatial increments. HRIRs are convolved with a source signal to create the illusion of directional sound over headphones. In a simple model of virtual Ambisonics, the pairs of HRIR convolutions required to render the spatialised scene (irrespective of the proliferation of sound sources) is equal to the number of virtual speakers desired for the implementation. Additionally, creation of head-tracked scene rendering is made considerably simpler in virtual Ambisonics by more streamlined rotation of the sound field, rather than continual convolution and interpolation between HRIRs for individual source signal locations [6].

1.1 Auditory display systems

In some contrast to the entertainment-focussed immersive media applications discussed above, virtual auditory displays (VAD) are typically multi-stream sonic environments that can be easily aurally segregated by users for information feedback or interaction purposes. To this end, there is a particular priority in spatialised VAD systems for maintaining clarity of sound source signals and accurate representation of their intended locations [7].

Recent analyses have shown that first order Ambisonics (FOA) provides inadequate representation of lateral and vertical localisation cues [8, 9, 10]. Currently, the only ways of improving the positional fidelity of B-Format decoding is either through higher order Ambisonics (HOA), or via more sophisticated decoding algorithms

[8]. Both HOA and alternate methods of B-Format decoding introduce mounting computational complexity and cost, which might not be achievable for the requirements of low-powered, portable real-time systems – i.e. embedded computing, wearable technology or even mainstream mobile platforms where only a defined proportion of CPU resource might be available for audio handling.

1.2 Virtual VBAP

VBAP is an alternate spatial rendering technique that extends standard amplitude panning. It places sound sources by triangulating outputs from loudspeaker arrays, the only restriction for which is that each speaker must be positioned at uniform distance from the listener. VBAP offers improved fidelity of source signal representation compared to Ambisonics, since the least number of speakers required to render a source is always used – i.e. one, two or three – thus minimising phase issues. Moreover, since VBAP positioning is achieved by simple gain weightings for each input signal to every speaker feed, it is a more efficient spatialisation technique to implement than Ambisonics [11]. Comparative evaluation using auditory modelling suggests that triplet-wise application of VBAP produces considerably more stable lateralisation than either 2D or 3D second order Ambisonics [12].

The principles used in virtual Ambisonics can also be applied to achieve VBAP rendering over headphones. A binaural implementation of VBAP has previously been outlined as the non-diffuse sound rendering component of the Directional Audio Coding (DirAC) spatial reproduction method [13]. However, to the authors' knowledge, the potential benefits, applications and challenges of implementing a pure virtual VBAP approach has not been prominently considered in existing research literature.

1.3 Considerations for a virtual VBAP approach

VBAP's flexibility towards loudspeaker array configuration means it is straightforward to place a higher concentration of virtual speakers towards front-facing positions, without any additional computational effort [11]. Doing so can potentially enable increased positional resolution for locations where we tend to experience greater acuity [14, 7]. Moreover, in a binaural context, head-tracking is usually applied by fixing virtual speakers relative to the listener's head position whilst the sound scene itself is rotated [15, 6]. Virtual VBAP could therefore enable a defined frontal zone of improved auditory focus, which shifts automatically with head motion to any point in the virtual scene.

However, VBAP renders each input signal as an individual point source. For binaural rendering with virtual loudspeakers, this means that every sound stream must be rotated to its head-tracked position, weighted and summed to relevant virtual speaker feeds before real-time convolution with HRIRs. Applying reverberation in this configuration is a more problematic task than in virtual Ambisonics, where 3D room simulation can be first generated synthetically before resulting reflections are encoded in the B-Format domain [6].

Reverberation is recognised as a critical component of binaural rendering. Evidence suggests that early reflections alone offer optimal conditions to improve localisation of azimuth (but at some expense to elevation perception) and increase sense of externalisation [16]. Recent research also indicates that anechoic binaural signals rendered with artificial stereo or even mono reverb could be a perceptually viable approach to room simulation. These conditions compared favourably to both first and third order Ambisonically generated surround reverbs when judged in terms of both overall realism and fidelity of source location [17]. For these reasons, we regard implementation of reverb in virtual VBAP a separate research question that should not prevent objective investigation of the method's performance here.

The next section outlines the hardware used for real-time audio processing and head-tracking, the software implementation of virtual VBAP and first order Ambisonics rendering systems and the methodology used for evaluation. Following this, the outcomes of the technical assessment in terms of localisation cue error in both approaches are presented. We then discuss the significance of these findings for real-time 3D binaural VAD.

2 Methods

Two real-time head-tracked binaural systems were developed on the same embedded platform: a virtual VBAP renderer and a virtual FOA renderer. The hardware used and software design are outlined in detail below, along with the methodology used for comparative evaluation.

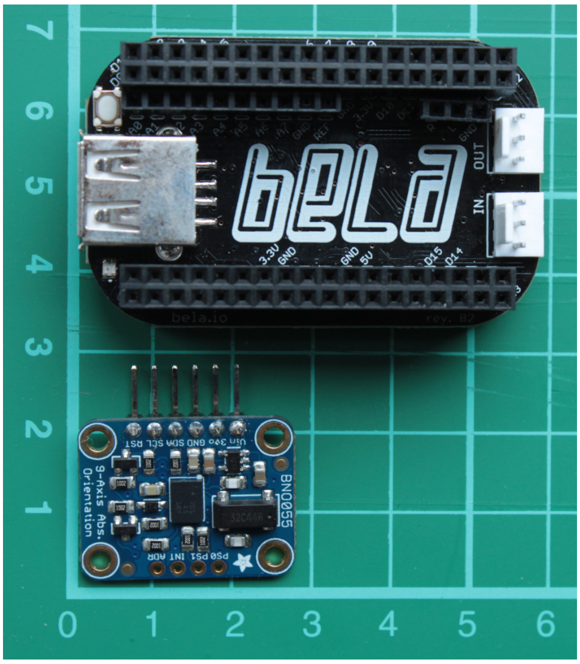

Fig. 1: The Bela Mini embedded Linux platform (top) and Bosch BNO055 IMU (bottom). The grid illustrates system dimensions (in centimetres).

2.1 Hardware

Bela is a commercially-available, open-source, embedded Linux platform for low-latency audio and sensor processing. The software described here was run on the compact "key-fob" sized Bela Mini model (Figure 1), built to achieve a high degree of portability. The Bela Mini runs a 1GHz ARM Cortex-A8 processor and has two channels of audio output [18]. Audio processing on the full sized Bela (which also runs a 1GHz ARM Cortex-A8) has been shown to perform at sub-millisecond levels of round-trip latency [19]. C++ was used to code both the VBAP and Ambisonics systems outlined below.

MrHeadTracker is a low-cost plug-and-play system for incorporating head-tracking on Arduino embedded computing platforms. It uses the Bosch BNO055 nine degrees of freedom sensor with on-board processing to measure, compute and output quarternions or Euler angles (Figure 1). The device performs with a refresh rate of 100Hz and is shown to have an angular standard deviation between 0.5° and 2.5° [20]. The fourth author ported the *MrHeadTracker* code for integration with Bela [21].

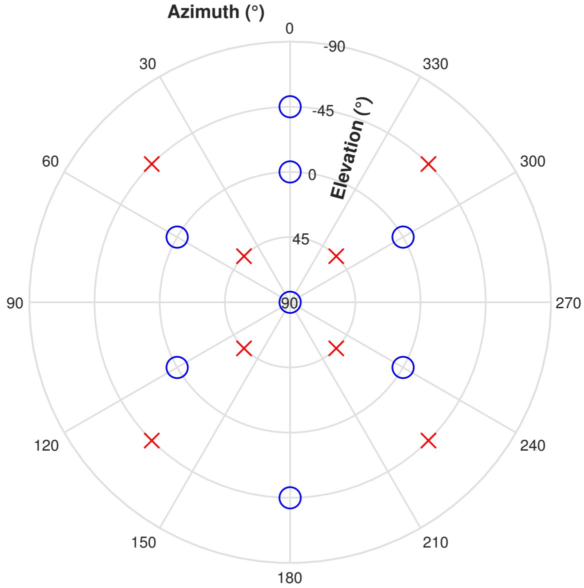

Fig. 2: Graphical representation of both systems' virtual speaker positions. **O** = VBAP system; **X** = FOA system.

2.2 VBAP binaural rendering software

Generating a binaural signal requires an HRTF set to synthesise the position of virtual speakers. The IRCAM LISTEN database has been chosen as the source from which to select a human-measured HRTF for this investigation [22]. This collection is favoured because it has been systematically rationalised into an optimal shortlist of seven that, collectively, is judged to present the highest strength of preference for the largest group of users [23]. (The usefulness of a system that allows user selection from a shortlist of human-measured HRTF sets is pertinent to ongoing and related research that is discussed more fully in [24].) LISTEN HRTF set 1013 was randomly chosen from the optimised shortlist for use in the subsequent analysis.

3D VBAP allows any combination of three or more loudspeakers placed equidistantly from the listener to render spatial audio. A working constraint of eight virtual speakers (therefore 16 individual convolutions with left and right HRIRs to render the binaural scene) has been established for realising the auditory environment. The aim is for this system to generate the effect of pseudo-spherical surround sound within that constraint. As described in the previous section, there is also an intent to concentrate more loudspeakers towards

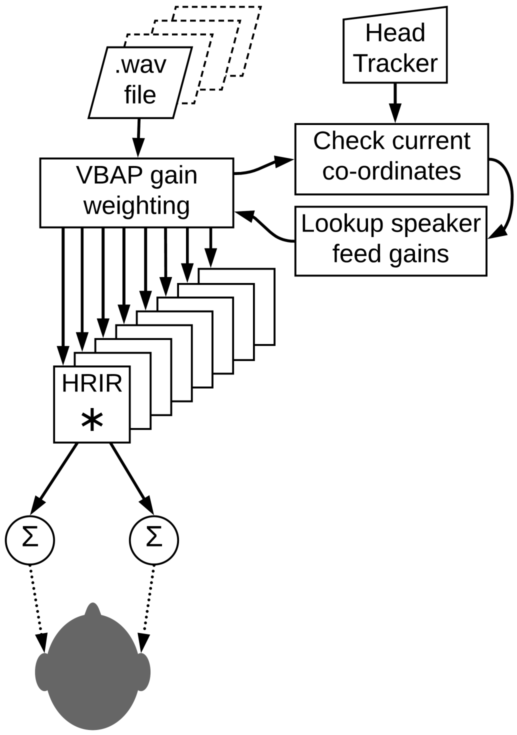

Fig. 3: Flow diagram of virtual VBAP implementation.

frontal azimuth locations for improved resolution. The position of each virtual speaker is given in Figure 2. The five placed on the horizontal plane (at 0° elevation) are spread in 60° increments, but with a 120° gap at the rear. A single speaker at the zenith enables upward triangulation. Since the HRTF set does not feature a measurement at the nadir (as is typically the case), two placed at -45° allow downward triangulation.

A flow diagram of the virtual VBAP system implementation is shown in Figure 3. VBAP weightings for the virtual loudspeaker layout have been precalculated using [25]. The resulting lookup table of speaker feed gains for every possible angular position at 1° increments is loaded as a matrix on startup. Standard resolution audio files (16 bit 44.1kHz) are called and streamed from any predefined azimuth/elevation coordinate. Head-tracking readings are refreshed at 86Hz (every 512 samples) to update the position of each sound source. A buffer of 2048 samples is used to comfortably meet processing deadlines for Fourier transformations and frequency domain convolution. The Bela digital-to-analogue converter is known to introduce 21 samples of delay [19]. Maximum system latency is

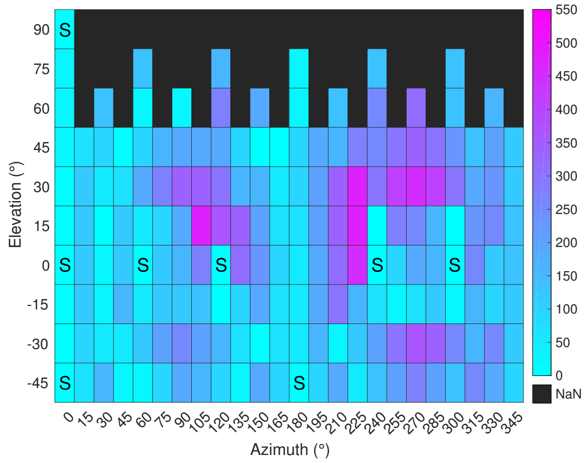

Fig. 4: Virtual VBAP unsigned ITD error (μs).
S = virtual speaker location.

therefore 2581 samples, or 59 milliseconds, which is within the 75 millisecond response time advocated by [26].

2.3 Ambisonics reference software

To evaluate the VBAP software's performance, the same hardware was used to render a virtual FOA environment. This system adapted the *libspatialaudio* C++ encoding/decoding library, which adopts MaxRe weighted and All-round Ambisonic Decoding algorithms for psychoacoustically optimised source representation [27]. Head-tracking was included by using BNO055 output data to rotate the B-Format sound field using on board functions. The library renders 3D FOA binaurally using virtual speakers placed in a cube arrangement with a chosen HRTF set. However, the processing required to run this code in its original form proved too computationally intensive for Bela. Instead, binauralisation was achieved using an Ambisonics-to-binaural optimisation, which calculates direct transfer functions for each B-Format channel, the principle for which is outlined in [15]. HRIRs from LISTEN 1013 were again used to generate binaural B-Format, which then only required eight convolutions to render 3D FOA at virtual speaker locations defined in Figure 2.

2.4 Localisation cue error measurement

Auditory perception of spatial location can be quantified by three measurements: interaural time difference

	All locations	Front locations*	-45° elevation	0° elevation	45° elevation
VBAP	150 (114)	115 (98)	120 (79)	140 (120)	151 (98)
FOA	202 (123)	175 (117)	157 (97)	216 (138)	246 (140)

Table 1: Mean and standard deviation of unsigned ITD error (μs) for virtual VBAP and FOA systems, by location

** Front locations include all elevation points within 0 to +/- 60° azimuth*

(ITD), interaural level difference (ILD) and monoaural spectral cues introduced by filtering from pinnae and (to lesser extents) head and torso reflections. Comprehensive discussion of these three cues' importance for spatial sound localisation is provided in [14, 7]. HRTFs encode into HRIRs the ITD, ILD and spectral shaping experienced by an individual at given spatial locations. In effect, both the VBAP and FOA systems synthesise new sets of HRIRs via virtual loudspeaker realisation. The response of these synthetic HRIRs can be compared directly to the original HRTF set from which they were derived and in respect of each of these localisation cues [10, 8]. To achieve this, unit impulse signals were fed into either system at each of the 187 LISTEN database co-ordinates and the outputs processed to derive localisation cue metrics compared to the original HRTF set, as follows:

- *ITD* – Measurement techniques advocated in [28] were applied to calculate ITD. VBAP-generated, FOA-generated and original HRIRs were first filtered with a tenth order Butterworth lowpass filter at 3kHz. ITDs for each set and position were then computed in microseconds using cross-correlation functions from the MATLAB Signal Processing Toolbox.

- *ILD* – VBAP-generated, FOA-generated and original HRIRs were first filtered with a tenth order Butterworth highpass filter at 1.5kHz. ILDs for each set and position were then computed as the mean-squared power difference in decibels.

- *Spectral response* – Peak normalisation was applied to both the VBAP-generated and the FOA-generated HRIRs. In each case and to ensure uniform gain increase, normalisation was referenced to the most significant unsigned value found in either the left or right channel of all 187 HRIRs viewed collectively. The VBAP-generated, FOA-generated and original HRIRs were then processed with a 40 band gammatone filter bank, with lower and upper centre frequencies at 0.1kHz and 16kHz [29]. For the VBAP-generated, FOA-generated and original HRIRs, spectral response was calculated as the mean-squared power of each band, for either channel, at every position.

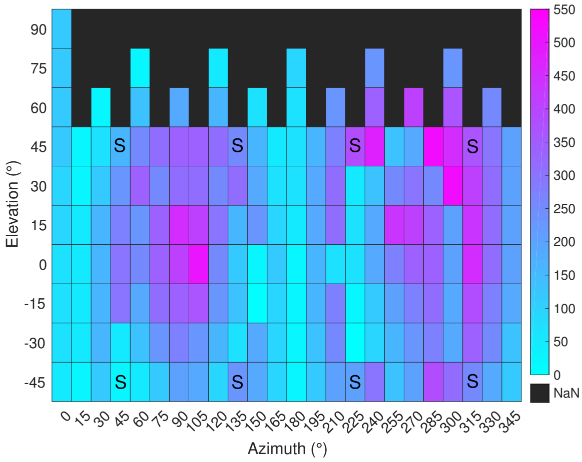

Fig. 5: Virtual FOA unsigned ITD error (μs). S = virtual speaker location.

3 Results

This section presents how the above described ITD, ILD and spectral shaping error measurements are used to evaluate the VBAP and FOA systems against the original LISTEN 1013 HRTF set ground truth.

3.1 ITD Error

Unsigned ITD errors for each system compared against the full LISTEN 1013 HRTF set are presented in Figures 4 and 5, for all 187 locations. As expected, no error is seen at positions where virtual VBAP speakers are located. At these points the signal is reproduced solely with the LISTEN 1013 HRIR for that origin.

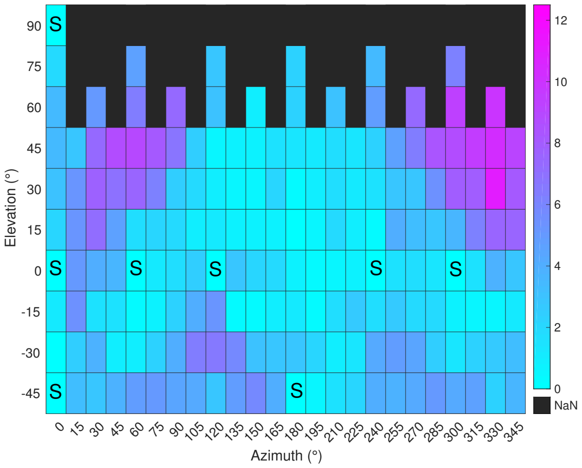

Fig. 6: Virtual VBAP unsigned ILD error (dB).
S = virtual speaker location.

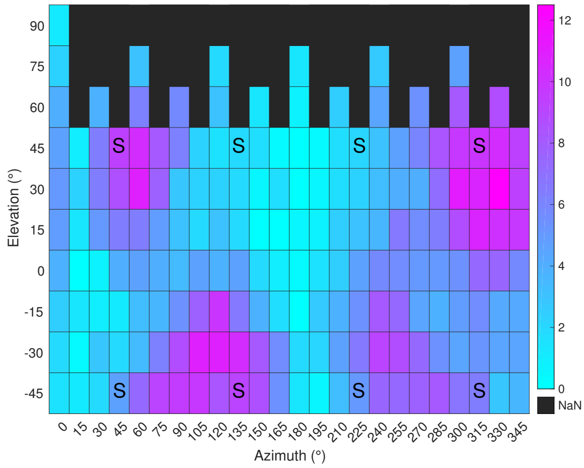

Fig. 7: Virtual FOA unsigned ILD error (dB).
S = virtual speaker location.

This is not the case with a FOA, where there is no clear relationship evident between ITD error and speaker location. Overall, greater deviation from the ground truth is generally apparent across wider areas in the case of FOA.

These tendencies can be seen more clearly in Table 1, where error rates are summarised into location groups. Inaccuracies are seen to be consistently lower and more stable in the VBAP implementation than FOA across every zone. In particular, virtual VBAP shows considerably less error than FOA at frontal locations and on the horizontal plane – areas where the VBAP speaker layout has been specifically concentrated to provide greater resolution. Even at locations where the VBAP speakers are most dispersed and FOA more concentrated (+ or - 45° elevation), virtual VBAP ITD is more closely aligned to the original HRTF set.

3.2 ILD Error

Unsigned ILD errors for each system when compared against the full LISTEN 1013 HRTF set are presented in Figures 6 and 7 for all 187 locations. In this instance, the pattern of error is similar between implementations, but its extent is greater for FOA. Virtual VBAP again benefits from points of no error at speaker locations, where FOA does not. Table 1 also confirms that VBAP outperforms FOA in every location grouping for ILD, with less than half the average error of FOA at frontal and horizontal plane positions.

3.3 Spectral Error

The mean unsigned spectral errors for each system and channel when compared against the full LISTEN 1013 HRTF set are presented in Figure 8. FOA has a lower mean spectral error for both left and right channels quite consistently up to about 1-2kHz. Above 2kHz, there is a fairly steady and relatively even increase in mean error seen for both systems and in either channel.

Figure 9 gives some illustration of where the error manifests at different locations. Points showing ≈ 0dB error occur at virtual VBAP speaker positions, as expected. It is also apparent that the FOA error, although generally lower, tends to show greater asymmetry between left and right channels than VBAP. This is particularly so at lower frequencies, where the performance of FOA is seemingly better in aggregate (in Figure 8). The VBAP system error is somewhat more symmetric between channels than FOA. There are slight indications that the VBAP virtual loudspeaker configuration has resulted in lower spectral error in some frontal locations (particularly at 0° elevation). However, unlike with ITD and ILD error, in general spectral response at frontal locations is not clearly shown to be any more in line with the original LISTEN HRIRs than FOA.

4 Discussion

Impulse response analysis suggests the virtual VBAP system has clear reproduction benefits for horizontal

	All locations	Front locations*	-45° elevation	0° elevation	45° elevation
VBAP	3.18 (2.58)	1.75 (1.45)	3.43 (1.61)	1.62 (1.43)	4.83 (3.49)
FOA	4.69 (3.05)	3.90 (2.50)	5.66 (3.03)	3.85 (2.08)	4.91 (3.69)

Table 2: Mean and standard deviation of unsigned ILD error (dB) for virtual VBAP and FOA systems, by location

** Front locations include all elevation points within 0 to +/- 60° azimuth*

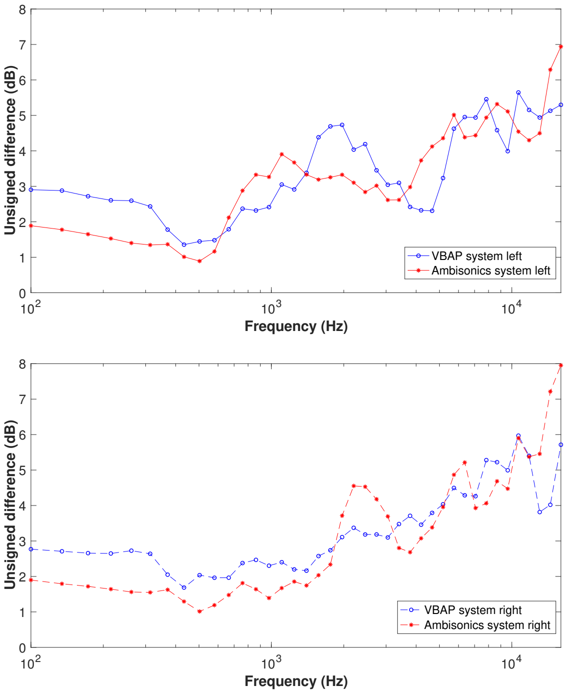

Fig. 8: Mean unsigned spectral error for all locations, by frequency band, system and channel

localisation. This configuration seems to reconstruct original HRIR ITDs with fewer "black spots" of significant divergence than FOA and with lower error overall. ITD is most pertinent to lateral perception of frequencies below 1.6kHz, thus it could be expected that VBAP would provide a more faithful representation of source azimuth for low-frequency dominated sounds and from a broader range of locations. Although the pattern of ILD error is loosely shared by both systems, VBAP again shows closer adherence to the original HRIRs across all positions. ILD provides the dominant horizontal cue above 1.6kHz, so it can be anticipated that localisation would also be more acute throughout the sound field within this frequency range[14].

In contrast, spectral error analysis suggests FOA has an aggregate frequency response that seems more faithful to source HRIRs. A simple interpretation might hypothesise that FOA could therefore present a more reliable representation of the original HRTF set's elevation cues, since frequency shaping inherent in HRIRs is key to vertical localisation. However, the majority of elevation cues are derived from HRIR frequency responses above 4kHz [8], where spectral error patterns between the two implementations become similarly discontinuous. As such, there is actually little data to show that HRTF set elevation cues would be preserved any more faithfully by the FOA system.

In fact, evidence of left/right ear asymmetries in the FOA spectral error (particularly in lower frequencies) suggests that extent of colouration is more location-dependent than for VBAP. It is possible that this unevenness is a byproduct of irregularities in the human-measured HRTF set itself – whether down to subject morphology or marginal discrepancies in the measurement procedure – in which case the VBAP configuration appears to smooth discontinuity with more even error distribution. Understanding the perceptual effect of these contrasting left/right channel spectral error patterns would require further investigation.

Finally, it is clear that on the horizontal plane VBAP rendering represents the source HRTF set's ITD, ILD and spectral response more closely. Having more faithful rendering on this particular plane is a significant benefit to any VAD system for source segregation purposes. Furthermore, frontal locations are also shown to reproduce HRTF set ITDs and ILDs more accurately with the proposed VBAP loudspeaker configuration. Combined with head-tracking, this approach enables dynamic enhanced resolution rendering across the 3D sound field.

5 Summary

We have presented a portable embedded system for rendering multi-channel spatial audio scenes binaurally

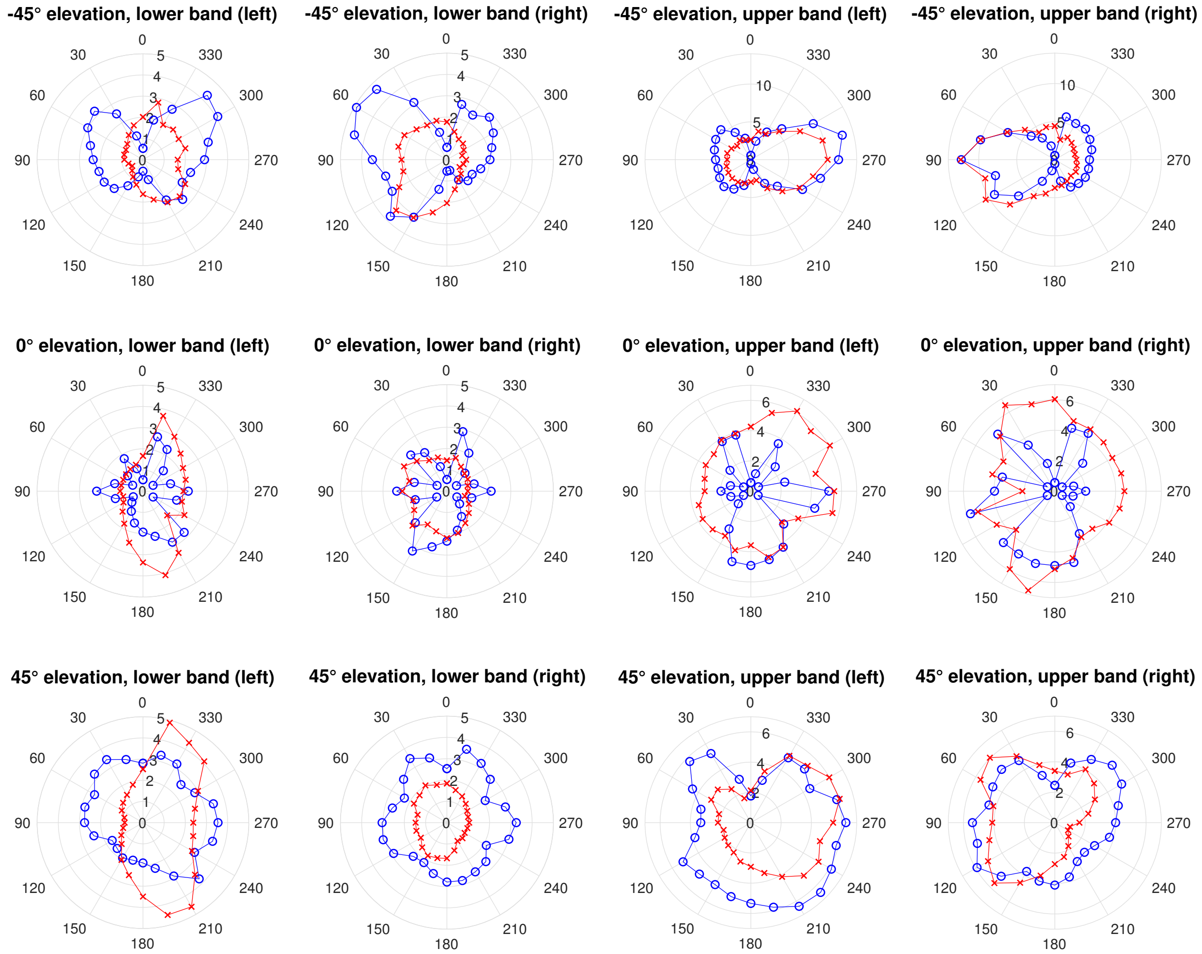

Fig. 9: Mean unsigned spectral error (dB) for selected locations and in upper and lower frequency band groupings**, by system and channel. **–o–** = VBAP system; **–x–** = FOA system. (Note differing scales to amplify plots).

*** The lower band grouping includes the first 17 filters, with centre frequencies ranging from 0.1 to 1kHz. The upper band grouping includes the following 23 filters, with centres at 1.5 to 16kHz*

in real time, using virtual VBAP. The system supports head-tracking within recommended levels of latency. Further development is underway to optimise the system with a view to incorporating user-selection of preferred HRTF sets, real-time sound source interaction and artificial reverberation. We develop the system with VAD research and practice in mind, including graphics-less user interfaces, multi-modal art installations or interactive live performance applications.

The technical analysis presented here shows that the system outperforms an equivalent FOA setup when assessed against key localisation metrics. These specific findings related to virtual VBAP are in line with previous investigations into FOA and VBAP spatial reproduction more generally. Indications from that existing research and the findings uncovered here suggest that second or even third order Ambisonics is required to meet localisation cue reproduction accuracy of virtual VBAP. This would require considerably more computational resource that might not be available on embedded devices or other working contexts with similar CPU resource constraints.

6 Acknowledgements

This work was supported by EPSRC and AHRC under the grant EP/L01632X/1 (Centre for Doctoral Training

in Media and Arts Technology).

References

[1] Google, "ResonanceAudio APIs," `https://developers.google.com/resonance-audio/reference/overview`, 2017.

[2] Facebook, "Spatial Workstation audio suite," `https://facebook360.fb.com/spatial-workstation/`, 2018.

[3] Unity, "Ambisonics format interfacing," `https://docs.unity3d.com/Manual/AmbisonicAudio.html`, 2018.

[4] Fellgett, P., "Ambisonics. Part one: General system description," *Studio Sound*, 17(40), pp. 20–22, 1975.

[5] Gerzon, M., "Ambisonics. Part two: Studio techniques," *Studio Sound*, 17(40), pp. 24–28, 1975.

[6] Noisternig, M., Musil, T., Sontacchi, A., and Höldrich, R., "3D binaural sound reproduction using a virtual ambisonic approach," in *VECIMS 2003 - 2003 International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, pp. 174–178, Lugano, Switzerland, 2003.

[7] Begault, D. R., *3D sound for virtual reality and multimedia*, Academic Press Limited, London, UK, 1 edition, 1994.

[8] Kearney, G. and Doyle, T., "Height perception in Ambisonic based binaural decoding," in *Audio Engineering Society Convention 139*, pp. 1–10, New York, NY, USA, 2015.

[9] Thresh, L., Armstrong, C., and Kearney, G., "A direct comparison of localisation performance when using first, third and fifth order Ambisonics for real loudspeaker and virtual loudspeaker rendering," in *Audio Engineering Society Convention 143*, pp. 1–9, New York, NY, USA, 2017.

[10] Wiggins, B., "Analysis of binaural cue matching using ambisonics to binaural decoding techniques," in *4th International Conference on Spatial Audio*, Graz, Austria, 2017.

[11] Pulkki, V., "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, 45(6), pp. 456–466, 1997.

[12] Pulkki, V., "Evaluating spatial sound with binaural auditory model," in *Proceedings of the International Computer Music Conference*, pp. 73–76, Michigan Publishing, Havana, Cuba, 2001.

[13] Laitinen, M. V. and Pulkki, V., "Binaural reproduction for directional audio coding," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 337–340, New Paltz, NY, USA, 2009.

[14] Blauert, J., *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

[15] McKeag, A. and McGrath, D. S., "Sound field format to binaural decoder with head tracking," in *Audio Engineering Society 6th Australian Regional Convention*, pp. 1–9, Melbourne, Australia, 1996.

[16] Begault, D. R., Wenzel, E. M., and Anderson, M. R., "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source." *Journal of the Audio Engineering Society*, 49(10), pp. 904–916, 2001.

[17] Picinali, L., Wallin, A., Levtov, Y., and Poirier-Quinot, D., "Comparative perceptual evaluation between different methods for implementing reverberation in a binaural context," in *Audio Engineering Society Convention 142*, pp. 1–7, Berlin, Germany, 2017.

[18] Bela, "Introducing the new Bela Mini," `https://blog.bela.io/2018/02/22/bela-mini-launch/`, 2018.

[19] McPherson, A., Jack, R., and Moro, G., "Action-Sound Latency: Are Our Tools Fast Enough?" in *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 16 of *2220-4806*, pp. 20–25, Queensland Conservatorium Griffith University, Brisbane, Australia, 2016.

[20] Romanov, M., Berghold, P., Frank, M., Rudrich, D., Zaunschirm, M., and Zotter, F., "Implementation and evaluation of a low-cost headtracker for

binaural synthesis," in *Audio Engineering Society Convention 142*, Berlin, Germany, 2017.

[21] Stewart, R., "Bela On Ur Head - head-tracking for binaural audio with Bela," https://github.com/theleadingzero/belaonurhead, 2018.

[22] Warusfel, O., "Listen HRTF database," http://recherche.ircam.fr/equipes/salles/listen/index.html, 2003.

[23] Katz, B. F. G. and Parseihian, G., "Perceptually based head-related transfer function database optimization," *The Journal of the Acoustical Society of America*, 131(2), pp. EL99–EL105, 2012.

[24] Shukla, R., Stewart, R., Roginska, A., and Sandler, M., "User Selection of Optiman HRTF Sets via Holistic Comparative Evaluation," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, pp. 1–10, Redmond, WA, USA, 2018.

[25] Politis, A., "Vector base amplitude panning library," https://github.com/polarch/Vector-Base-Amplitude-Panning, 2015.

[26] Suzuki, Y., Yairi, S., and Iwaya, Y., "Effect of large system latency of virtual auditory display on listener's head movement in sound localization task," in *Proceeding of the 13th International Conference on Auditory Display*, pp. 24–31, Montréal, Canada, 2007.

[27] Digenis, A., "Ambisonic encoding / decoding and binauralization library in C++," https://github.com/videolabs/libspatialaudio, 2018.

[28] Katz, B. F. G. and Noisternig, M., "A comparative study of interaural time delay estimation methods," *The Journal of the Acoustical Society of America*, 135(6), pp. 3530–3540, 2014.

[29] IoSR Surrey, "General purpose Matlab toolbox," https://github.com/IoSR-Surrey/MatlabToolbox, 2017.