3D TIME-BASED AURAL DATA REPRESENTATION USING D^4 LIBRARY'S LAYER BASED AMPLITUDE PANNING ALGORITHM

Ivica Ico Bukvic

Virginia Tech
SOPA, DISIS, ICAT
Blacksburg, VA, USA
ico@vt.edu

ABSTRACT

The following paper introduces a new Layer Based Amplitude Panning algorithm and supporting D^4 library of rapid prototyping tools for the 3D time-based data representation using sound. The algorithm is designed to scale and support a broad array of configurations, with particular focus on High Density Loudspeaker Arrays (HDLAs). The supporting rapid prototyping tools are designed to leverage oculocentric strategies to importing, editing, and rendering data, offering an array of innovative approaches to spatial data editing and representation through the use of sound in HDLA scenarios. The ensuing D^4 ecosystem aims to address the shortcomings of existing approaches to spatial aural representation of data, offers unique opportunities for furthering research in the spatial data audification and sonification, as well as transportable and scalable spatial media creation and production.

1. INTRODUCTION

In todays rich data driven society strategies for optimal data experience and comprehension are more important than ever. Humans are biologically predisposed to experiencing rich environmental data multimodally [1], warranting research into individual modalities' potential in promoting data comprehension and interpretation delivered through technology. Such research serves as the foundation for their combined utilization to broaden cognitive bandwidth and clarity [2]. In this respect, visual data exploration or visualization has arguably seen greatest progress. This may be in part because of human predisposition to visual stimuli, as well as because visualizations have had a rich history [3] that both predates and inspires todays technology-centric approaches.

Audification and sonification [4] are relatively new but nonetheless thriving research areas. In particular, they offer a diverse array of complementing and competing approaches to spatial aural representation of data. With auditory spatial awareness covering practically all directions [5], it is a dimension that exceeds the perceivable spatial range of the visual domain. Apart from the simple amplitude panning [6], audio spatialization approaches include Ambisonics [7], Binaural [8], Depth Based Amplitude Panning (DBAP) [9], Vector Based Amplitude Panning (VBAP) [10], and Wave Field Synthesis (WFS) [11]. The following paper focuses primarily on spatialization strategies that are reproducible in physical environments and offer physical affordances with minimal amount of idiosyncrasies, such as the vantage point, without requiring additional technological support, e.g. a motion tracking system. For this reason, due to its specific context that does not meet the aforesaid criteria the paper excludes the Binaural approach from the discussion below.

2. CATALYST

This project was inspired primarily by the newfound space whose hybrid HDLA implementation exposed new audio spatialization research opportunities and challenges. Virginia Tech Institute for Creativity, Arts, and Technology's (ICAT) Cube is an innovative space with a hybrid audio infrastructure capable of supporting all of the aforesaid approaches to spatializing sound, with particular focus on WFS, Ambisonics, and VBAP (Fig.1). It is a 50x40x32-foot (WxLxH) blackbox space with catwalks and mesh ceiling whose audio infrastructure is centered around the idea of discovery and experimentation, including audification and sonification. ICAT's Cube offers a unique hybrid 148-channel audio system designed in collaboration with ARUP Acoustics inc. In order to accommodate the various spatialization algorithms, it consists of a 124.4 homogeneous loudspeaker array offering several horizontal layers of varying density: a high density ear-level equidistant 64-channel array and additional three loudspeaker layers with 20 channels each, including a 20-channel ceiling raster. The 124-channel system is complemented by 4 symmetrically positioned subs centered on each side of the first level catwalk. The system also offers an additional 17-inch sub focusing primarily on sub-50Hz frequencies. It can be further complemented by 10 mobile floor-level loudspeakers. Cube also offers nine ceiling-mounted ultrasonic audio spotlights, including four mounted onto a motorized, remotely-controlled arm.

In a space designed for transdisciplinary research that needs to be capable of near seamlessly transitioning from one spatialization technique to another and/or concurrently employing multiple approaches, such an implementation is not without a compromise. Cube's WFS relies on a proprietary Sonic Emotion Wave 1 system [12] that enables its implementation using sparser loudspeaker configuration. Ambisonics require careful calibration due to cuboid shape of the loudspeaker configuration [10]. Finally, VBAP due to algorithm's inability to handle irregular densities, particularly the ear-level layer, utilizes only select ear-level and ceiling loudspeakers, therefore relying more on the virtual sound positioning than what a localized amplitude panning system may

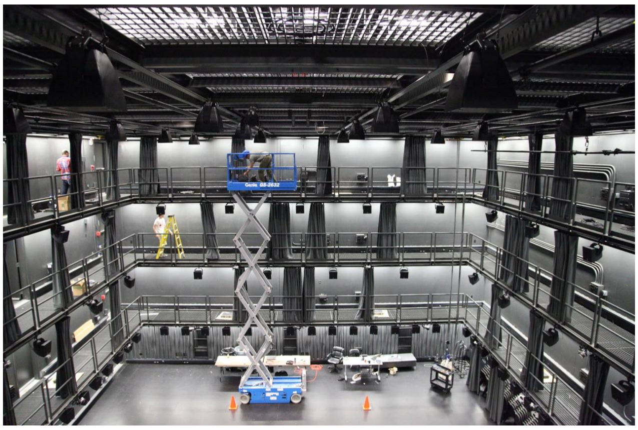

Figure 1: Virginia Tech ICAT Cube.

ostensibly require (Fig.3). Furthermore, each of the aforesaid configurations provides limited transportability among various spaces, including 900 square foot ICAT Perform Studio's 2-layer 24.4 Genelec system, and the the Digital Interactive Sound & Intermedia Studio (DISIS) classroom offering 8.2 single-layer Genelec system. Apart from the WFS plane wave [11] and VBAP's source spread (a.k.a. MDAP) [13] that manifests itself in a form of a regular circle-like shape around the source's center, none of the spatialization approaches offer an easy and controlled way of projecting sounds through multiple physical sources, particularly when it comes to irregular shapes. Likewise, none of the currently available technologies provide the aforesaid features in a way that can easily scale among varying loudspeaker configurations while utilizing all of the available physical sources and their superior localization over that of virtual ones [14].

Based on the observations attained through the newfound Virginia Tech signature audio research space, several inconsistencies have emerged that limit broader applicability of the preexisting approaches to audio spatialization with particular focus on audification and sonification scenarios (listed in no particular order):

1. Support for irregular High Density Loudspeaker Arrays (HDLAs);
2. Focus on the ground truth with minimal amount of idiosyncrasies;
3. Leverage vantage point to promote data comprehension;
4. Optimized, lean, scalable, and accessible, and
5. Ease of use through supporting rapid-prototyping time-based tools.

Recent Computer Music Journal solicitation [15] has defined HDLAs as "systems addressing 24 or more independent loudspeaker". In this paper HDLAs are further defined as loudspeaker configurations capable of rendering 3D sound without having to rely solely on virtual sources or post-processing techniques.

2.1. Support for flexible loudspeaker layouts

While most of the aforesaid spatialization algorithms are HDLA and therefore 3D capable, most implementations favor certain loudspeaker configurations, e.g. tightly spaced loudspeakers in WFS or triangular loudspeaker placement in VBAP and High-Order Ambisonics [16]. DBAP is configuration-agnostic, but also requires additional features, such as *spatial blur*, designed to minimize problems of spatial coloration and spread typical of both VBAP an DBAP [9]. Recent research further suggests for some of these approaches there may be ways to utilize less common configurations (e.g. Blue Ripple Sound's Rapture3D for irregular loudspeaker arrangements using HOA [17] or the proprietary Sonic Emotion systems that allow for sparse WFS arrays [12]). Due to their proprietary nature, currently the limits of these solutions is not known, nor how well and/or how reliably they may be able to scale and/or accommodate systems whose irregularity significantly deviates from the prescribed configuration (e.g. sparse vs. irregular loudspeaker distribution in WFS), particularly in HDLA environments. VBAP solution, like the one implemented in the Virginia Tech's Cube utilizes only some of the loudspeakers in order to attain the desired triangular organization among the loudspeakers, leaving a number of physical sources unused (Fig.3). Given the physical sources' superior audio spatialization potential over the virtual ones, such a solution was found incapable of harnessing the full potential of CUBE's audio system.

2.2. Ground Truth with Minimal Amount of Idiosyncrasies

Each of the aforesaid spatialization approaches is encumbered by unique idiosyncrasies that limit the ease of their applicability in a broad range of scenarios. These idioscyncrasies can be seen as a detriment towards developing generalizable sonification strategies in part because they can also cloud the prospect of identifying the ground truth. The most obvious one is the aforesaid sensitivity of various approaches to loudspeaker configurations. Similarly, positioning a virtual (e.g. Ambisonics [7] and specialized cases in WFS [18]), and physical sound sources (e.g. 4DSound [19]) inside the listening area offers great promise. Yet, their respective idiosyncrasies, such as the sweet spot (e.g. WFS aliasing, and lower order Ambisonics), custom and ostensibly intrusive hardware (4DSound), and the computational complexity (e.g. Ambisoncs) that currently lacks out-of-box solutions, particularly when associated with non-standard loudspeaker layouts, limits their universal applicability. Similarly, WFS's ability to place sounds outside the listening space may allow for more uniform perception of the sound source, yet doing so will also limit the power of the vantage point that may help in clarifying source's location and its relationship to other adjacent sources depending on listener's location. Lastly, DBAP introduces spatial blur to compensate for potential spatial coloration and spread inconsistencies.

2.3. Leveraging Vantage Point

In this paper the author posits for a system to provide optimal listening environment, it needs to mimic affordances of our everyday lives as long as its implementation does not exacerbate one of the observed limitations. Vantage point is one such affordance that enables listeners to perceive both the rendered aural data within the context of their immediate environment, as well as perceive rendered data communally in a location-specific fashion. Unlike virtual sources within the listening area that also introduce limiting idiosyncrasies, vantage point is essentially intrinsic to simpler amplitude-based algorithms. This allows for a closer study of a particular angle, or even positioning oneself closer to the loudspeaker perimeter to elevate perceived amplitude of a source or texture of interest, something that may prove particularly useful in data audification and sonification.

The vantage point limitation brings out another important consideration in pursuing a more universal and transportable approach to spatializing sound–loudspeaker perimeter based spatialization. Instead of relying on idiosyncratic sound processing that enables rendering of virtual sources within and outside the loudspeaker perimeter, the spatialization should ideally focus on the perimeter-centric rendition, an approach that offers relatively straightforward mapping of multidimensional data onto the loudspeaker perimeter, reinforces the vantage point, and makes it considerably easier to reproduce in varied and flexible HDLA scenarios.

2.4. Optimized, Lean, Scalable, and Accessible

Ideally, a system should be lean–it should rely on the preexisting tool frameworks where possible, ensuring that at its very core it is simple and maintainable with minimal redundancies. This is certainly the case with some of the implementations that are typically embedded in digital signal processing languages, including Max [20], Pure-Data [21], and Supercollider [22], or provided as plugins (e.g. VST, LV2 plugins, or Audio Units). Such implementations can leverage the vast resources of those toolkits to further enhance their functionality and flexibility.

In terms of rendering spatial data using sound, one of the additional considerations is system's responsiveness and how that responsiveness scales from conventional stereophonic to HDLA scenarios. Ideally, such a system should be capable of rendering a scene in real-time and under low-latency conditions. While low-latency operation is not necessarily critical in controlled tests, its absence may limit system's applicability and broader appeal, both of which are essential for wider adoption and potential standardization across multiple sites and contexts.

Although all of the aforesaid systems offer real-time and low-latency performance, some (e.g. WFS and HOA) require careful space- and loudspeaker-layout-specific calibration that may not be easily accessible out-of-box. In particular, when considering systems with cutting-edge features (e.g. Wave 1), their proprietary nature may render them as prohibitively expensive black box implementations with more complex HDLA configurations requiring special design and licensing. This can also be seen as a potential factor in limiting the access to such solutions and consequently their transportability.

2.5. Rapid-Prototyping Tools

If implemented well, rapid prototyping tools have a unique ability to go well beyond representing loudspeaker positions and their respective amplitudes. By interfacing with multidimensional data sources, such tools have the potential to lead to cross-pollination of generalizable standards across various modalities and by doing so serve as a scaffolding in domains whose standards are yet to be solidified. For instance, being able to interact with visual representation of audio spatialization may lead towards leveraging standards and techniques associated with visual drawing and painting and using those to guide the development of corresponding methodologies in the spatial aural domain.

Sound is a time-based modality and for this reason, rapid prototyping tools should go beyond providing the ability to position a sound source. They could also include a way of altering their location over time, as well as visualizing the outcomes of such a change. With the exception of Sonic Emotion's Wave 1 [12], 4DSound [19], D-Mitri system [23], VBAP-based Zirkonium [24], and recently introduced Sound Particles [25], HDLA spatialization systems are devoid of any time-based data that can be easily synced with other time-based content (e.g. video or an abstract data feed), typically requiring users to create their own middleware to drive such systems in real-time and/or render their audio feeds into a multichannel audio file. While offering ability to visualize loudspeaker configuration, it is currently unclear if DBAP offers any rapid prototyping tools. Similarly, it remains unclear whether Sound Particles is capable of rendering real-time low-latency audio nor what is its CPU overhead in doing so.

Within the context of audification and sonification, none of the existing off-the-shelf systems offer easy interfacing with multidimensional data sets and their translation into a spatialized sound.

2.6. Other Considerations

Based on the observed limitations, the author of this paper posits that the ideal platform for pursuing a generalizable approach to spatial data representation using sound should mimic as closely as possible real-world environmental conditions our multisensory mechanisms are accustomed to experiencing, leveraging, and interpreting. More so, it should do so with minimal technological complexity and idiosyncratic limitations. Such a system is more likely to integrate and cross-pollinate with other modalities and in return leverage their preexisting body of research to identify optimal mapping strategies. Furthermore, the author argues that such cross-pollination in particular between visual and aural may offer a useful scaffolding to sonification theory based on the existing body of research in the visual domain. In a pursuit of such a solution the technology presented in this paper focuses primarily on data sets with up to four dimensions.

3. INTRODUCING D^4

D^4 is a new Max [20] spatialization library that aims to address the aforesaid limitations by:

1. Introducing a new lean, transportable, and scalable audio spatialization algorithm capable of scaling from monophonic to HLDA environments, with particular focus on advanced spatial manipulations of sound in audification and sonification scenarios, and
2. Providing a collection of supporting rapid prototyping time-based tools that leverage the newfound audio spatialization algorithm and enable users to efficiently design and deploy complex spatial audio images.

Below we will focus primarily on the spatialization algorithm that in part builds on author's prior research [26] and its newfound affordances that have a potential to serve as a foundation for the further exploration of the auditory display paradigm.

3.1. D^4's Algorithm

At the very core, D^4 is driven by the newly proposed Layer Based Amplitude Panning (LBAP) algorithm. LBAP is rooted in a straightforward sinusoidal amplitude panning algorithm which amounts to:

$$L_{\text{amp}} = cos(L_{\text{distance}} * \pi/2), \quad (1)$$

$$R_{\text{amp}} = sin(L_{\text{distance}} * \pi/2), \quad (2)$$

L and R variables stand for left and right channels spatially oriented from listener's perspective in clockwise fashion, respectively. L_{distance} is a normalized value between 0 and 1 and where the ensuing amplitude value between 0 and 1 is used to modulate the outgoing audio signal for both L and R channels.

In 2D arrays of varying densities, e.g. horizontal ear-level arrays, the math for manipulation between loudspeakers remains essentially the same, with the only addition being the awareness of the loudspeaker and source positions in horizontal space expressed as an angle (0-360 degrees). By a simple calculation, one can either identify perfect physical source (a loudspeaker) or two adjacent loudspeakers and using the aforesaid function calculate the amplitude ratios between the two. What makes this approach particularly convenient is its ability to utilize irregular densities across the perimeter with the only caveat being decreased angle perception resolution in areas that may be sparser in terms of loudspeaker spacing and therefore more reliant on virtual sources (Fig.2).

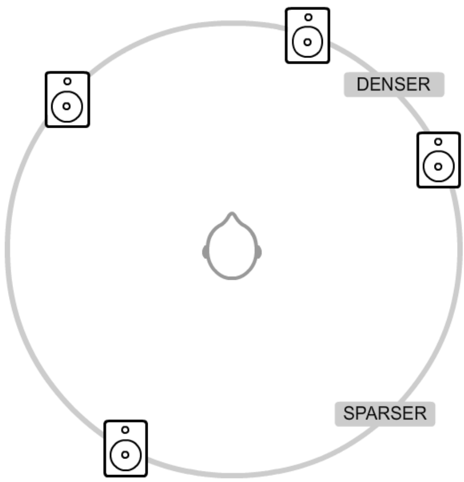

Figure 2: Irregular 2D perimeter loudspeaker array's localized reliance on virtual sources and its inversely proportional relationship to the array's immediate density.

When applying the same algorithm in a 3D environment where there are multiple horizontal layers of loudspeakers positioned around the perimeter, the aforesaid algorithm is typically superseded by VBAP [27] and more recently DBAP [9]. Where VBAP begins to fall apart is when horizontal loudspeaker layers are populated with varying densities and consequently irregular distances among loudspeakers. This is certainly the case with the ICAT Cube where the upper levels host only 20 loudspeakers as opposed to 64 loudspeakers at ear level and where such a configuration make sense given human decreased spatial perception accuracy of elevated sound sources. This, however, is not the only scenario. Similar limitations can theoretically also occur in spaces whose architectural design precludes equal loudspeaker distribution, something that DBAP aims to address albeit with added complexity and ensuing idiosyncrasies. For instance, there may be acoustic considerations, structural beams, pillars, walls, and other physical structures that prevent loudspeaker placement. When employing VBAP, such setups fail to provide usable adjacent triangles, as is the case with ICAT Cube (Fig.3), and while one can skip physical sources in order to retain triangular configuration, such a solution precludes the use of all physical sources, resulting in a less than ideal scenario, particularly when considering preferred higher loudspeaker density at ear level where human perception, depending on head orientation, offers greatest angular resolution. Another option is using a hybrid system, so that the secondary spatialization approach utilizes the higher density layer. This, however, further limits system's transportability and introduces an entirely new array of idiosyncrasies.

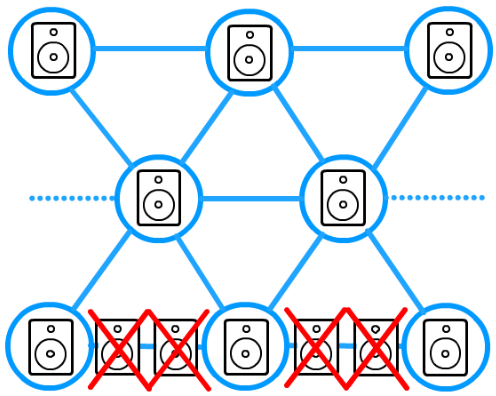

Figure 3: VBAP's selective use of loudspeakers in irregular layered loudspeaker configurations.

LBAP aims to address this problem by introducing an amplitude panning variant that relies on the core notion that the entire perimeter-based audio system is separated into a series of layers with each layer being assigned shared elevation and each loudspeaker further identified by its azimuth (Fig.4). In this respect vertical surfaces with loudspeaker rasters above, as is the case with the ICAT Cube's ceiling, and below are treated as a series of concentric circles, which is a feature that loosely resembles D-mitri and MIAP's grouping. In cases where there are no loudspeakers below or above, the lowest and highest layers assume any sound that moves below or above their elevation respectively should be cross-faded across the layer itself (Fig.5). While this is less than ideal, it can be easily remedied by adding an additional layer below (should the architecture allow for doing so), while leveraging existing infrastructure to the best of its ability.

Once the layers are identified, LBAP uses one vertical cross-section as the elevation reference. Doing so will enable for the sound to easily traverse individual layers horizontally (as it should) without having to compensate for vantage point deviations in elevation (e.g. loudspeaker in a far corner will effectively have lower elevation than one immediately next to the listener that belongs to the same layer (Fig.6). In cases where sound does not neatly fall onto one of the physical sources or a single horizontal layer, LBAP based on source's elevation first identifies its closest two layers, the one below and one above where the virtual source is located. Once the two layers are identified LBAP calculates their respective amplitude ratios as follows:

$$Above_{\text{amp}} = cos(Below_{\text{distance}} * \pi/2), \quad (3)$$

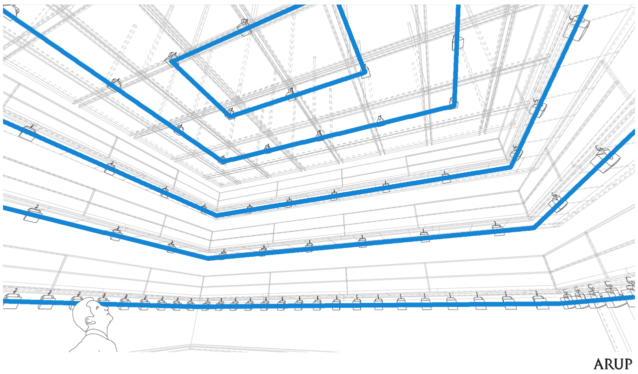

Figure 4: ICAT Cube's HDLA split into layers, including the ceiling raster. Space render courtesy of ARUP Inc.

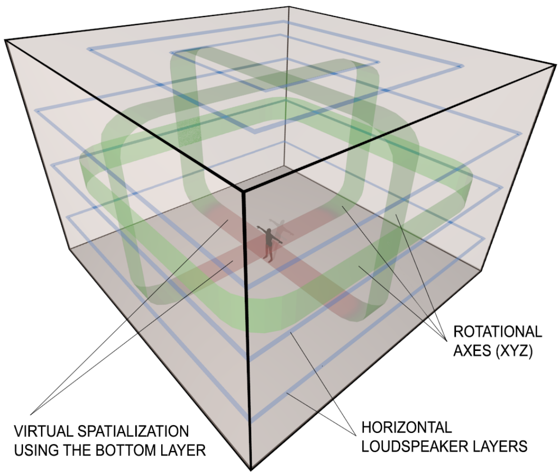

Figure 5: Vertical source rotations in ICAT Cube's environment below ear level rely entirely on virtual sources due to lack of physical layers.

$$Below_{\text{amp}} = sin(Below_{\text{distance}} * \pi/2), \quad (4)$$

The layer elevation is expressed in degrees from -90 to 90, which when combined with azimuth allows for describing all possible angles. Above refers to the layer above, and Below to the layer below the source's position. $\text{Below}_{\text{distance}}$ refers to the distance in degrees from the lower layer normalized so that the full distance between the two layers is equal to 1. The resulting layer amplitudes are calculated using the sinusoidal amplitude panning approach. The layer amplitude values are then used to modulate the output amplitude of the neighboring loudspeakers whose amplitude values have been calculated based on source's azimuth using the same sinusoidal approach:

Below layer:

$$BL_{\text{amp}} = cos(BL_{\text{distance}} * \pi/2) * cos(Below_{\text{amp}} * \pi/2), \quad (5)$$

$$BR_{\text{amp}} = sin(BL_{\text{distance}} * \pi/2) * cos(Below_{\text{amp}} * \pi/2), \quad (6)$$

Above layer:

$$AL_{\text{amp}} = cos(AL_{\text{distance}} * \pi/2) * cos(Above_{\text{amp}} * \pi/2), \quad (7)$$

$$AR_{\text{amp}} = sin(AL_{\text{distance}} * \pi/2) * cos(Above_{\text{amp}} * \pi/2), \quad (8)$$

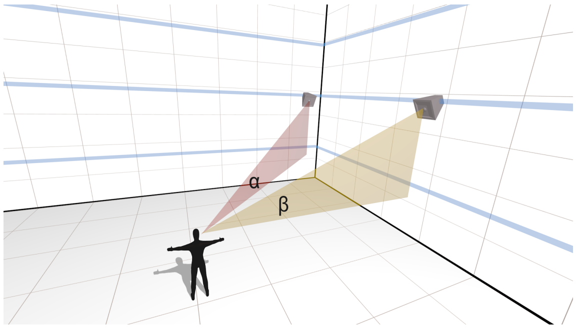

Figure 6: In a layered approach, depending on the architecture, loudspeakers within the same layer may have slight angle anomalies from the perceiver's vantage point, as is the case here with angles α and β.

The ensuing two-step amplitude panning algorithm variant is effectively loudspeaker density agnostic. One layer can have a few loudspeakers, while other many. Regardless of the configuration, the algorithm will never utilize more than four loudspeakers for point sources. It is important to emphasize the layer elevation that is calculated using vertical cross-section of the space will undoubtedly deviate for other loudspeakers in the space based on listener's position. Given, LBAP treats 3D loudspeaker arrangements as perimeter-based spatial canvas, such deviations are seen as being within the tolerance range of human perception, as they effectively mimic limitations of cinematic screens where certain aspects of the image from an individual vantage point are closer or farther, resulting in seemingly illogical proportions, yet in our minds we assemble such an image as a whole by taking into account their relative relationships. Similarly, in informal listening tests, LBAP has proven capable of rendering horizontally moving sounds that were higher than ear-level while still projecting a sense of horizontal, rather than vertically erratic motion due to vantage point variances in individual loudspeaker elevation within a particular layer.

3.1.1. Moving Sources

Once a point audio source is placed in a location, it can be rotated horizontally using azimuth and vertically using elevation, with the assumption it always emanates from the perimeter. The special case for spherically moving sound sources are situations where due to lack of additional physical layers (e.g. in the case of the ICAT Cube there are by default no layers lower than the ear level) the sound may have to be panned across the space, inferring sound at that point is being panned inside the listening area, rather than above or below the listener, something the system lacking physical sources is clearly incapable of rendering convincingly. This, however, is primarily a hardware limitation and is for the most part spatialization algorithm agnostic.

3.1.2. Independent Layers

Given sub channels are often treated as a separate group, spatializing sources based on their own layered design, D^4 allows for defining layers whose amplitude computation takes into account each such layer independently. This has proven instrumental in its integration into the ICAT Cube which utilizes four subs centered on each of the four sides of the 1st level catwalk as the first independent layer and with an additional 17-inch subwoofer that provides rumbling lows for the entire space from a single source as the second independent layer. Consequently, the algorithm inherently allows for use of a single loudspeaker per layer, resulting in 100% of the original generated amplitude emanating from that speaker regardless of the source's position and/or radius.

3.2. Advanced Sonification Features

Apart from WFS' plane wave [11] or VBAP's source spread (a.k.a. MDAP) [13] that manifests itself in a form of a circle-like shape around the source's center projected onto the 3D loudspeaker perimeter, none of the spatialization approaches offer an easy and controlled way of projecting sounds through multiple physical sources, particularly when it comes to irregular shapes. What arguably sets D^4 apart from other spatialization algorithms is its re-invisioned approach to growing point sources using Radius, and the Spatial Mask, as well as a suite of supporting spatialization tools that leverage these newfound affordances.

3.2.1. Radius

Each point source's default radius is assumed to be 1°. As it grows, based on proximity calculated as a linear distance between source's location and radius and physical loudspeaker's position, it spills over adjacent loudspeakers with its amplitude decreasing in all directions using the sinusoidal amplitude panning curve. As a result sounds with a diameter of 180°, cover entire sphere with the opposite edge being essentially inaudible. At 360° diameter, the overlap between the outer diameters when coupled together (and further limited not to exceed maximum allowable amplitude) amount to 100% of the original amplitude (Fig.7).

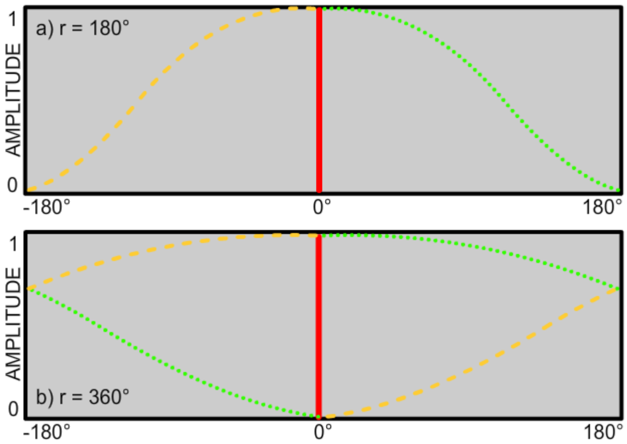

Figure 7: Sinusoidal amplitude curve applied to source radius in a single dimension. Thick red line denotes relative source's angle within one dimension. Yellow striped and green dotted lines denote two radius vectors across the said dimension. Example a) shows 180° diameter or 90° radius with no overlap, and b) 360° diameter or 180° radius with overlap.

3.2.2. Spatialization Mask

D^4's Spatial Mask (SM), akin to that of its visual counterpart considers the entire spherical space to have the default mask of 1. This means wherever the point source and whatever its radius, it will populate all the loudspeakers based on the computed amplitude. The spatial mask, however, can be changed with its default resolution down to 0.5° horizontally and 1° vertically, giving each loudspeaker a unique maximum possible amplitude as a float point value between 0 and 1. As a result, a moving source's amplitude will be limited by its corresponding mask value as it traverses the ensuing spherical perimeter. This also allows a situation where a point source with 180° radius or 360° diameter that emanates throughout all the loudspeakers can now be dynamically modified to map to any possible mask. When coupled with time-based visual editing tools, this equates to essentially aural painting [26] in both 2D and 3D. We will further explore SM and its features as part of the D^4 Rapid Prototyping Tools section below.

4. SIDE-STEPPING LIMITATIONS

D^4's implementation of the LBAP algorithm is a lean implementation in that it relies on Max's framework. Consequently, when coupled with Max's battery of digital signal processing objects, it allows for greater extensibility. For instance, through the use of a collection of included abstractions, D^4 library offers access to an otherwise complex form of movable sound sources, including angled circular motion, and the ability to control attack and trail envelope, effectively resulting in the aural equivalent of the motion blur (MB). D^4 also offers easy way of interfacing with Max's Jitter library [28] that offers vector optimization when calculating multidimensional matrices, something that has proven particularly useful when working with the Spatialization Mask.

Informal LBAP an D^4 tests have shown it is capable of providing critical low-latency real-time rendering of spatialized audio sources even in ¿100 HDLA and high-audio-stream-count scenarios. This makes it particularly useful in interactive environments. Its inaugural implementation as part of a Tornado simulation that premiered in the fall 2014 features 1,011 internal 24-bit 48KHz audio streams or channels stemming from two dozen concurrent point sources that are mixed down and outputted through the 124.4 CUBE loudspeaker system in real-time with audio latency of 11ms (512-byte buffer) between the time an action is initiated and the sound leaving the computer. D^4's implementation of the LBAP algorithm is designed to scale from monaural to as many loudspeakers as the system (CPU and audio hardware) can support. The current version offers a growing array of optimizations, including omission of unnecessary audio streams and bypassing redundant requests.

D^4 offers both single- and multi-threaded implementations. The multithreaded version, however, has offered only marginal improvement over its single-threaded counterpart. This is likely due to the fact that the built-in algorithm's implementation maximizes reliance on the built-in Max objects and as such in and of itself does not bear significant CPU footprint. More so, whatever the savings in terms of CPU utilization due to distribution across multiple CPU cores are replaced by the newfound overhead required to synchronize concurrent audio streams through a high number of interrupts required by the low-latency setting. Further testing is warranted to attain a better understanding of the CPU overhead in single- and multi-threaded scenarios.

One of the greatest challenges of the HDLA audio content is its transportability. Fixed media tends to be distributed as pre-rendered multichannel sound files that are often accompanied by a simple Max patch or an equivalent tool capable of interfacing with often unconventional HDLA configurations. The target venue, however, may not have the same number of loudspeakers, requiring either sound to be re-rendered (assuming the existing system lends itself to easy reconfiguration), or calling for compromises in determining which channels need to be omitted or doubled. As an alternative, a live version may be used where sound sources are coupled by a system that renders entire piece in real-time, requiring engine that is adaptable and reconfigurable. D^4 aims to address transportability by providing a simple one-step reconfiguration consisting of loudspeaker channels and their respective azimuths and elevations provided in an ordered (bottom-up, clockwise) layered approach that instantly updates all instances within the Max ecosystem and adapts the spatialization algorithm for a newfound loudspeaker arrangement. With its real-time low-latency scalable engine D^4 can also leverage the aforesaid implementation within the live and interactive aural spatialization of data, as well as artistic contexts.

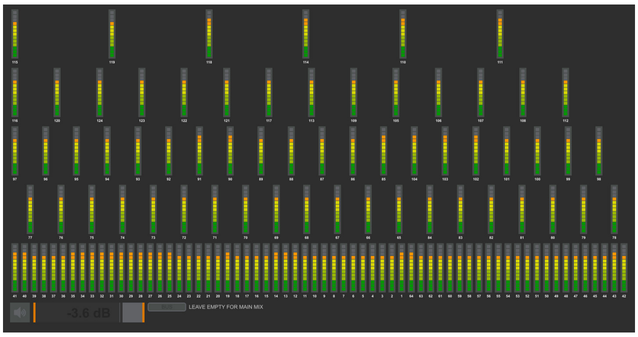

Figure 8: An instance of D^4 library's signal monitor.

With its integration into Max, D^4 immediately benefits from the built-in debugging and signal monitoring tools. With the help of the Jitter library, its spatialization capabilities can be easily translated into visual domain. The same has enabled D^4 access to external control surfaces. For instance, the aforesaid Tornado simulation offers iPad interface for controlling the simulation from the Cube's floor through the use of the Max's Mira library. In addition, it offers a visual level monitor built out of a collection of abstractions that enable users to easily customize and design new space-specific level monitors. Given the exponential complexity of signal flow in HDLA scenarios, the entire D^4 ecosystem is virtual audio bus aware, and offers a collection of visual tools, a.k.a. monitors specially tailored to harness this feature (Fig.8). By assigning a bus name to a particular monitor, it will automatically switch to monitoring all outputs from that bus, while leaving the bus name blank will revert to monitoring main outs. Similarly, the library provides a global main out whose adjustments affect all its instances. By default, D^4 comes with monitors for three Virginia Tech spaces, including DISIS, and ICAT's Cube and Perform studio, and offers easy way of creating new site-specific level monitors using a collection of abstractions.

5. D^4'S RAPID PROTOTYPING TOOLS

D^4 library also offers a series of rapid prototyping tools. Below we'll provide a brief overview of its 2D and 3D editors and means of importing data sets.

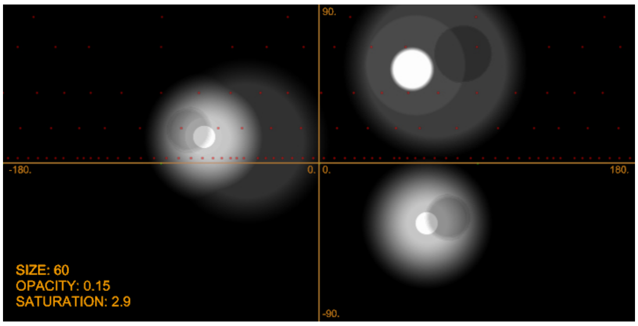

Figure 9: D^4 library's 2D mask editor.

2D mask editor (Fig.9) is a two-dimensional representation of the loudspeaker perimeter unfolded onto a plane with the x axis covering the full circle and y axis covering 90 degrees above and below. Apart from the usual representation of angles and a cursor, the visualization also auto-populates various layers, or as is the case with the ICAT Cube, the 124-speaker array (red dots) and its complementing 4-sub array (green dots). Jitter is used to allocate area around each loudspeaker up to the half-point between it and the adjacent loudspeakers. This area is used to calculate loudspeaker's overall amplitude based on its average grayscale color, with the black color denoting silence and white color 100% of the original amplitude.

To edit sound's mask, user is provided a customizable cursor that, akin to that of a digital drawing software, can be resized and its brush altered by varying transparency and saturation. Furthermore, the user can translate the SM both in conjunction with sound's rotation or independently of it. The editor also provides brush mirroring around the texture's x axis edges to simplify cross-fading across the visual seam generated by unfolding the mask onto a finite 2D plane. The ensuing mask can be fed either in real-time or on demand to the desired sound object. It can be also stored for time-based use and/or storage we will briefly discuss as part of the time-based editing features below.

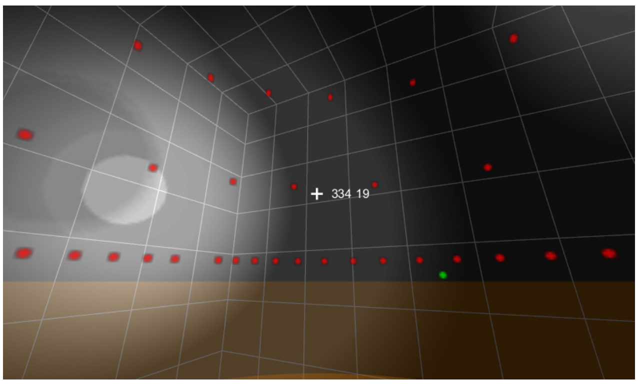

Figure 10: D^4 library's 3D mask editor.

3D Mask Editor (Fig.10) is a three-dimensional counterpart to the aforesaid 2D version. It allows for the exact same viewing and editing of the mask albeit from a 3D vantage point within the space, allowing user to pan the view around. This allows users to place 2D drawings in the context of the actual 3D space. While the default visualization provides a space-agnostic cuboid space, inspired by ICAT Cube's setup, given Max's flexibility, its layout can be easily altered, including importing actual 3D meshes of the target space. The ensuing drawing is stored in the identical way as the 2D drawing and the two are mutually interchangeable.

5.1. Time-Based Editing

While Mask Editors provide an easy way of populating sound sources in a cloud-like configuration throughout the HDLA space, their full potential is realized by leveraging accompanying time-based editor. Each of the mask snapshots can be stored as a 4-dimensional matrix (x, y, color, sequential keyframes). The matrix is accompanied by a coll object data structure which further contains timing of each keyframe and information on whether the transition between the current and next keyframe requires interpolation or not. In its initial release there is only one linear form of interpolation available between frames with other forms to be introduced in a later version based on user feedback. In addition, SM can be translated across x and y axes (corresponding to azimuth and elevation). Such interpolation is processed in parallel to interpolation (cross-fading) between SM keyframes. This can effectively serve as a secondary means of simulating cloud (as opposed to point) source's location.

Given the mask editor is fairly CPU intensive, for more complex real-time rendering saved editor renditions packaged as matrix-coll data containers can be retrieved and replayed using a considerably leaner Spatial Mask Player. Doing so enables playing multiple concurrent instances with minimal CPU overhead.

What makes D^4's approach to spatialization potentially useful as a platform for auditory displays is its ability to interface with the vast collections of spatial data and their translation into the 3D aural domain. Transferring visual data can be achieved by exporting it into an array of grayscale images, using format such as MJPEG, and importing it as a matrix into the editor. By relying on this feature alone, one could separate RGB channels into separate layers, effectively creating an audification engine of a movie footage. D^4 editor also allows for synchronization with external clocks using SMPTE, and can adjust internal pace in respect to the sync.

6. ADVANTAGES

Based on the observed features, LBAP and the D^4 library offer a number of advantages over the existing approaches that may be relevant to the audio display research, as well as the live and production scenarios, including support for irregular HDLAs, transportability, focus on the ground truth with minimal idiosyncrasies, vantage-point aware, optimized, lean, scalable, and accessible, and with the help of a growing number of rapid prototyping tools, the ease of use with particular focus on mapping multidimensional data onto spatial audio.

D^4's design focuses on rapid prototyping and implementation, leveraging existing battery of Max objects wherever possible, and consequently the pursuit of maximum flexibility. Such hybrid, mostly open source (MOSS) approach to software distribution is envisioned to isolate aspects that are easily modifiable by community and thereby encourage iterative improvement through community participation, while retaining control over the core algorithm and its still evolving APIs. As a result, the library is also implemented as a potential drop-in replacement for the existing approaches to spatialization that predominantly rely on the azimuth/elevation value pairs. Although amplitude overages are unlikely, as a safety precaution, LBAP further implements hard limiting per physical output channel, preventing amplitudes that exceed 1 or 100% of the incoming sound.

7. LIMITATIONS AND FUTURE WORK

While LBAP's simplicity essentially makes it capable of addressing just about any 3D loudspeaker layout that can be reasonably described as a collection of horizontal layers, D^4 library's rapid prototyping tools do not account for corners in cuboid scenarios as potentially special cases, something that would affect both the amplitude and the vantage point elevation. In informal listening tests the dichotomy between the assumed spherical azimuth/elevation loudspeaker location assignment and the actual cuboid layout of the ICAT Cube has not revealed observable deviations mainly because the azimuth and elevation hold true in both cases, with the ostensible amplitude variation due to differing distances between the listener and individual loudspeakers being below the observable threshold.

The same layered approach may make LBAP not applicable to certain scenarios. While some such scenarios are delineated for instance in DBAP paper [9], it is currently unclear how necessary or useful such a feature may be, particularly within the context of spatial audification and sonification. To address this, LBAP's layers could be ostensibly applied in a way where such layers are not treated as parallel, albeit at a potentially significant increase in algorithm's complexity.

Unlike D-Mitri and MIAP, D^4 is currently not capable of grouping loudspeakers. While similar results can be achieved through the use of the Sound Mask and/or independent layers, there is clearly a need for potential use of groups in the system's future iterations. Another limitation is the lack of multiple independent multilayered contexts. Currently, the system supports one multilayered context and virtually unlimited number of additional independent layers. It is unclear whether it makes sense to have multiple concurrent multilayered contexts exist within the same space.

It is worth noting that Jitter operations D^4's SM relies on are not designed to take place per audio sample and as such its visual tools have more limited resolution than the audio itself. While the system provides built-in audio interpolation this remains one of the potential limitations, particularly when it comes to exploring innovative approaches to spatial amplitude modulation that goes beyond the fifty keyframes per second. Outside such extreme cases, the number of possible keyframes has proven more than adequate.

8. OBTAINING D^4

D^4 is currently under development with the anticipated commercial release in the summer 2016. For licensing enquiries contact the author at ico@vt.edu.

9. REFERENCES

[1] M. H. Giard and F. Peronnet, "Auditory-visual integration during multimodal object recognition in humans: a

behavioral and electrophysiological study," *Journal of cognitive neuroscience*, vol. 11, no. 5, pp. 473–490, 1999. [Online]. Available: http://www.mitpressjournals.org/doi/abs/10.1162/089892999563544

[2] S. Y. Mousavi, R. Low, and J. Sweller, "Reducing cognitive load by mixing auditory and visual presentation modes," *Journal of Educational Psychology*, vol. 87, no. 2, pp. 319–334, 1995.

[3] M. Friendly, "A Brief History of Data Visualization," in *Handbook of Data Visualization*, ser. Springer Handbooks Comp.Statistics. Springer Berlin Heidelberg, 2008, pp. 15–56, dOI: 10.1007/978-3-540-33037-0_2. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-33037-0_2

[4] G. Kramer, *Auditory display: Sonification, audification, and auditory interfaces*. Perseus Publishing, 1993. [Online]. Available: http://dl.acm.org/citation.cfm?id=529229

[5] T. R. Letowski and S. T. Letowski, "Auditory spatial perception: Auditory localization," DTIC Document, Tech. Rep., 2012. [Online]. Available: http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA563540

[6] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7853

[7] J. Daniel and S. Moreau, "Further study of sound field coding with higher order ambisonics," in *Audio Engineering Society Convention 116*. Audio Engineering Society, 2004. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12789

[8] B. Carty and V. Lazzarini, "Binaural HRTF based spatialisation: New approaches and implementation," in *DAFx 09 proceedings of the 12th International Conference on Digital Audio Effects, Politecnico di Milano, Como Campus, Sept. 1-4, Como, Italy*. Dept. of Electronic Engineering, Queen Mary Univ. of London,, 2009, pp. 1–6. [Online]. Available: http://eprints.maynoothuniversity.ie/2334

[9] T. Lossius, P. Baltazar, and T. de la Hogue, "DBAPdistance-based amplitude panning." Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2009. [Online]. Available: http://www.trondlossius.no/system/fileattachments/30/original/icmc2009-dbap-rev1.pdf

[10] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7853

[11] K. Brandenburg, S. Brix, and T. Sporer, "Wave field synthesis," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009*. IEEE, 2009, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5069680

[12] E. Corteel, A. Damien, and C. Ihssen, "Spatial sound reinforcement using Wave Field Synthesis. A case study at the Institut du Monde Arabe," in *27th TonmeisterTagung-VDT International Convention*, 2012. [Online]. Available: http://www.wfs-sound.com/wp-content/uploads/2015/03/TMT2012_CorteelEtAl_IMA_121124.pdf

[13] V. Pulkki, "Uniform spreading of amplitude panned virtual sources," in *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*. IEEE, 1999, pp. 187–190. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=810881

[14] G. Parseihian, C. Jouffrais, and B. F. Katz, "Reaching nearby sources: comparison between real and virtual sound and visual targets," *Frontiers in neuroscience*, vol. 8, 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4151089/

[15] "CMJ Call for Works." [Online]. Available: http://www.computermusicjournal.org/HDLA-call.html

[16] F. Hollerweger, "An Introduction to Higher Order Ambisonic ," *April 2005*, 2013. [Online]. Available: http://flo.mur.at/writings/HOA-intro.pdf

[17] "Custom Layouts in Rapture3d Advanced | Blue Ripple Sound." [Online]. Available: http://www.blueripplesound.com/custom-layouts

[18] E. Corteel, C. Kuhn-Rahloff, and R. Pellegrini, "Wave field synthesis rendering with increased aliasing frequency," in *Audio Engineering Society Convention 124*. Audio Engineering Society, 2008. [Online]. Available: http://www.aes.org/e-lib/online/browse.cfm?elib=14492

[19] "4dsound," http://4dsound.net/. [Online]. Available: http://4dsound.net/

[20] M. Puckette, "Max at seventeen," *Computer Music Journal*, vol. 26, no. 4, pp. 31–43, 2002. [Online]. Available: http://www.mitpressjournals.org/doi/pdf/10.1162/014892602320991356

[21] ——, "Pure Data: another integrated computer music environment," *IN PROCEEDINGS, INTERNATIONAL COMPUTER MUSIC CONFERENCE*, pp. 37–41, 1996. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.3903

[22] S. Wilson, D. Cottle, and N. Collins, *The SuperCollider Book*. The MIT Press, 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=2016692

[23] "D-Mitri : Digital Audio Platform | Meyer Sound," http://www.meyersound.com/product/d-mitri/spacemap.htm. [Online]. Available: http://www.meyersound.com/product/d-mitri/spacemap.htm

[24] C. Ramakrishnan, "Zirkonium: Non-invasive software for sound spatialisation," *Organised Sound*, vol. 14, no. 03, pp. 268–276, Dec. 2009. [Online]. Available: http://journals.cambridge.org/article_S1355771809990082

[25] "Sound Particles - Home," http://www.sound-particles.com/. [Online]. Available: http://www.sound-particles.com/

[26] I. Bukvic, D. Gracanin, and F. Quek, "Investigating artistic potential of the DREAM interface: The Aural Painting," in *Proceedings of the International Computer Music Conference*, 2008.

[27] J. Cofino, A. Barreto, and M. Adjouadi, "Comparing Two Methods of Sound Spatialization: Vector-Based Amplitude Panning (VBAP) Versus Linear Panning (LP)," in *Innovations and Advances in Computer, Information, Systems Sciences, and Engineering*, ser. Lecture Notes

in Electrical Engineering, K. Elleithy and T. Sobh, Eds. Springer New York, 2013, no. 152, pp. 359–370, dOI: 10.1007/978-1-4614-3535-8_31. [Online]. Available: http://link.springer.com/chapter/10.1007/978-1-4614-3535-8_31

[28] M. Song, S. A. Mokhov, and S. P. Mudur, "Course: Rapid advanced multimodal multi-device interactive application prototyping with Max/Jitter, processing, and OpenGL," in *Games Entertainment Media Conference (GEM), 2015 IEEE*. IEEE, 2015, pp. 1–2. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7377246