AMBISONICS: INTRODUCING AN ACCESSIBLE, DISTRIBUTED APPROACH

Thomas A. RUSHTON(thomas.rushton@inria.fr)[1], **Romain MICHON**[1], **Tanguy RISSET**[1], **Clemens WEGENER**[1], and **Pierre LECOMTE**[2]

[1] **Inria, INSA Lyon**, CITI, EA3720, 69621 Villeurbanne, France
[2] **École Centrale de Lyon**, 36 Av. Guy de Collongue, 69130 Écully, France

ABSTRACT

This paper presents an accessible, distributed approach to higher order ambisonics using a network of low-cost microcontroller-based devices. Leveraging recent advances in embedded computing, networked audio, and precise time synchronization, the system offers an alternative to conventional centralized architectures that typically require expensive multichannel interfaces and high-performance digital signal processing hardware. The proposed architecture combines an audio plugin for soundfield design with a modular network of ambisonics processing units, which, exploiting the parallelisable nature of ambisonics decoding, are each responsible for a subset of loudspeaker outputs. Synchronization is achieved through the IEEE 1588 Precision Time Protocol, complemented by adaptive clock correction to maintain sub-microsecond alignment between devices. A third-order ambisonics implementation demonstrates the viability of the approach, with potential scalability to higher orders and larger arrays. Overall, this work contributes to the development of frugal, distributed solutions for immersive audio, expanding access to ambisonics-based systems for a wide range of applications.

1. INTRODUCTION

Large-scale spatial and immersive audio systems remain exclusive commodities. The cost of loudspeakers aside (though various frugal and DIY solutions have been proposed [1, 2]), supporting hardware — highly multichannel audio interfaces and powerful, centralised DSP systems — is, for many researchers, artists, and other interested parties, prohibitively expensive; the barriers to entry to implementing wave field synthesis (WFS) [3] and higher order ambisonics (HOA) [4] systems remain stubbornly high.

Research in recent years has, however, witnessed efforts to lower these barriers to entry. Exploiting advancements in embedded and distributed computing, networked audio, plus a general spirit of frugality and DIY-mentality, experimental and low-cost spatial audio systems have been developed in attempts to provide viable alternatives to the commercial status quo [1, 2, 5–10].

Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

In this paper we present a distributed HOA system consisting of a digital audio workstation (DAW) plugin and a synchronous network of low-cost microcontroller (MCU) based devices. Exploiting the parallelisable nature of the ambisonics decoding process, the system is scalable and modular, with a minimum of redundancy, and provides a flexible framework for spatial audio reproduction, while highlighting remaining challenges in configuration and channel scalability. In principle any parallelisable spatial audio algorithm could be implemented on our system; here, in section 3, we give the example of a third order HOA installation, adding, in section 4, that it is potentially scalable to at least fifth order.

2. BACKGROUND

A variety of comparatively recent technological advancements have made the present work possible. These include the rise of embedded computing and *internet of things* platforms, the adoption of networked audio as the standard in multichannel audio transmission in local area settings, the introduction of the IEEE 1588 Precision Time Protocol (PTP), and much research in spatial and immersive audio. In the following subsections we describe the most pertinent contributing developments.

2.1 Frugal Spatial Audio

Recent years have seen various attempts to democratise and render spatial and immersive audio installations more accessible. These include frugal wave field synthesis arrays driven by a Field Programmable Gate Array (FPGA) [1, 9] and the authors' previous work on distributed sound field synthesis systems based on networked MCUs [7, 8, 10]. Due to the parallelisable nature of sound field synthesis algorithms, and their adeptness in parallel computation, the FPGA-based approach shows promise, but it is encumbered by the relative cost of FPGA development boards, plus the difficulty associated with programming them. High level tools for programming FPGAs, such as the SyFaLa toolchain [11], do exist, but nonetheless the complexity associated with FPGA development poses at the very least a steep learning curve.

The audio DIY and maker communities have benefited in the past decade or so from the emergence of a variety of *embedded* platforms for audio development, such as MCU-based devices — ESP32 and STM boards, plus the Daisy and Teensy — and embedded Linux systems — the Bela, Elk, and Raspberry Pi amongst the latter [12]. These

platforms offer a variety of audio-specific advantages: suites of audio-specific libraries in the case of the Daisy, Teensy and Bela, programmability in the high-level audio domain specific language (DSL) Faust in those and many other such cases [12], very low latency in the case of the Bela platform, and all-in-all a low-cost, small form-factor way of getting sophisticated audio processing and control interfaces into novel digital musical instruments and installations.

2.2 Distributed Spatial Audio

A fundamental accessibility issue by which spatial audio systems are afflicted is that of their typically centralised nature. In order to cover a large listening area with optimal spatial resolution, a large number of loudspeakers must be employed, which in turn traditionally entails the deployment of high channel-count audio interfaces. Even a comparatively small-scale WFS system may demand as many as 64 output channels [13]; a fifth-order HOA system calls for a minimum of $(5 + 1)^2 = 36$ channels, and ideally, to maximise the system's sweet-spot listening area, many more. Recent years have augured in efforts to *distribute* the problem of spatial audio, with both HOA [5] and WFS [6–8, 10] given due attention.

Distributed audio processing gives rise to inevitable problems around synchronisation. No two devices configured to run at a given sampling rate will run at exactly the same precise rate due to the manufacturing tolerance of their internal time reference, a crystal oscillator. It has been demonstrated that MCU-based devices possessing the appropriate functionality can be synchronised to sub-microsecond order via PTP [14, 15], and furthermore, that PTP-based time exchange can be used for media clock and presentation time recovery [10].

Despite being memory-poor, sporting only 1024 KB RAM [10], a Teensy 4.1 MCU development board equipped with audio and ethernet add-ons costs less than € 50 [10], and memory can be extended by up to 16 MB by way of inexpensive surface-mount PSRAM chips. [1] Furthermore, its NXP IMXRT1062 MCU [16] features an ARM core processor clocked at 600 MHz, more powerful than many similar MCU-based platforms [8]. Critically, for purposes of synchronisation in a distributed setting, the IMXRT1062 also features a PTP timer, hardware timestamping functionality, and a dedicated audio subsystem with extremely high-resolution phase-locked loop (PLL) numerator and divider registers — in effect it is possible to enact nanosecond-order adjustments to the sampling rate of the device during runtime to ensure synchronous reproduction across networked modules; indeed, although commercial multichannel audio interfaces exist at a pricepoint of € 150 for ten line outputs (e.g. Behringer UMC1820), if using multiple such interfaces for high channel-count applications there is no way to synchronise the devices. Furthermore, being programmable in Faust, the Teensy can benefit from the existence of the *AmbiTools* suite of Faust-based utilities for ambisonics [17].

[1] https://www.pjrc.com/store/psram.html

2.3 Ambisonics

Ambisonics is a sound field synthesis technique based on the principle of decomposing a sound field onto a truncated basis of spherical harmonics [4]; the *order* M of the basis truncation gives the total number of spherical harmonics used to decompose the sound field: $(M+1)^2$. Using a limited number of functions for the decomposition limits the spatial zone where the sound field can be accurately represented and reconstructed. Its size is such that $\lceil kr \rceil = M$, where r is the radius of the zone of accurate reconstruction (the *"sweet spot"*) and $k = 2\pi f/c$ is the wave number, f is frequency, and c the speed of sound.

First order ambisonics, for which $M = 1$, uses four signals, and was first theorized and developed by Gerzon [4]. HOA is implemented when $M > 1$. For sound field reconstruction the number of loudspeakers required is $N \geq (M + 1)^2$. Estimation of the ambisonic components of a sound pressure field is possible via recording with spherical microphone arrays [18]; alternatively, monophonic sound sources can be spatialised via *encoding* to ambisonic format [17], which, by convention, uses *ambisonics channel numbering* (ACN) to designate channels.

ACN channels are indexed by the degree m and order n of the Legendre polynomials by which they are described: [17]

$$a = m^2 + m + n \ . \tag{1}$$

For a single monophonic source $x(t)$ at azimuth θ and elevation φ, the encoded signal b_a for a given ACN-format channel is

$$b_a(t) = x(t) \cdot Y_{m,n}(\theta, \varphi) \ , \tag{2}$$

where $Y_{m,n}$ is the real valued spherical harmonic [17]

$$Y_{m,n}(\theta, \varphi) = N_{m,n} P_{m,|n|} \sin(\varphi) \begin{cases} \cos(|n|\theta), & n \geq 0 \\ \sin(|n|\theta), & n < 0 \end{cases} \ . \tag{3}$$

Here $N_{m,n}$ is a normalisation factor, and $P_{m,|n|}$ are the associated Legendre polynomials. Ambisonics is not without its terminological ambiguities [19]: it is the *degree* of the polynomial that corresponds with the ambisonics *order*.

Considering the case of a third order ambisonics system with the bare-minimum $(3 + 1)^2 = 16$ loudspeakers at positions θ_i, φ_i the spherical harmonics make up the 16×16 mode matrix $\mathbf{Y}_{m,n}(\theta, \varphi)$. In this case, the decoding matrix $\mathbf{D}$ is simply the inverse, $\mathbf{D} = \mathbf{Y}^{-1}$, the rows of which are the decoding coefficients for each loudspeaker; for a greater number of loudspeakers it would be necessary to take the pseudoinverse, $\mathbf{D} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T$.

The time-domain output signal for the i^{th} loudspeaker is found as the dot product of the i^{th} row of $\mathbf{D}$ and the 16×1 vector of ACN-format channels:

$$s_i(t) = \mathbf{D}_i \cdot \mathbf{b}(t) \ , \tag{4}$$

and it is clear, then, that the ambisonics decoder is parallelisable — a DSP module producing output for speakers i and j need only be aware of the coefficients for the corresponding i^{th} and j^{th} rows of the decoding matrix.

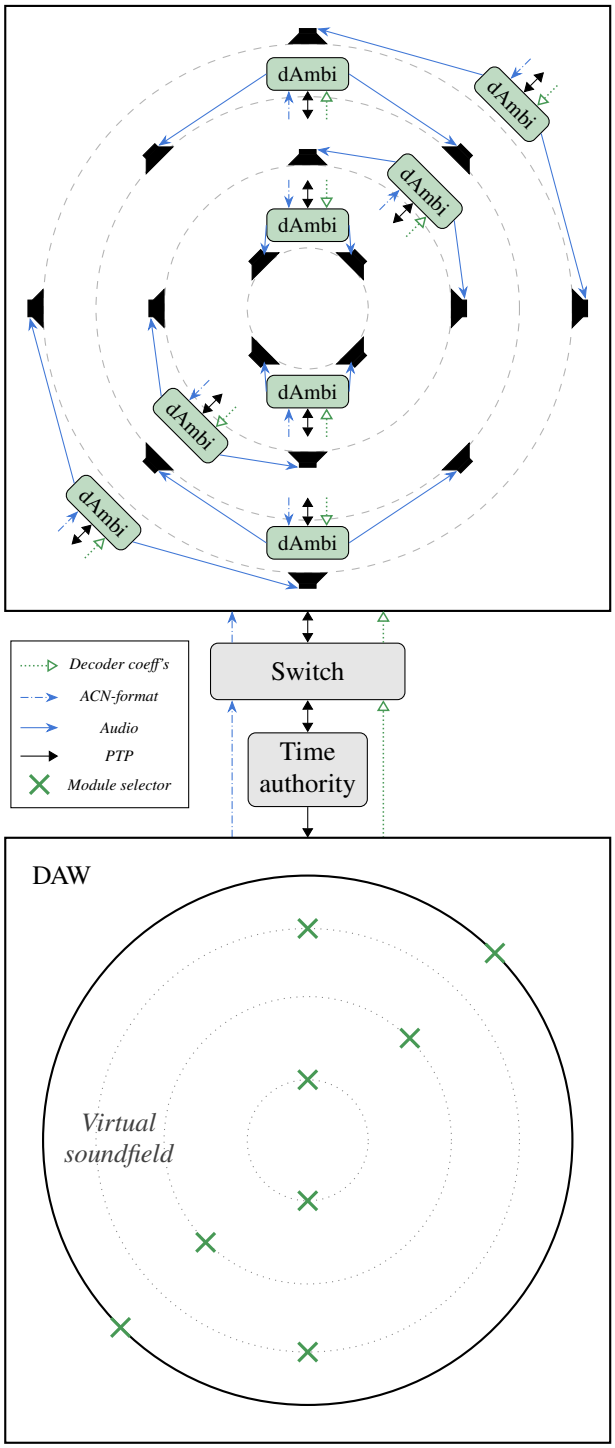


Figure 1. Example third order distributed ambisonics setup. A DAW plugin describes a virtual soundfield and permits the selection of distributed ambisonics (*dAmbi*) modules. Decoding coefficients and ACN-format encoded audio are transmitted via the network switch. The time authority provides a temporal frame of reference for the system as a whole.

3. IMPLEMENTATION

Our system consists of a DAW plugin featuring an interface for describing a virtual sound field, coupled with a network of distributed Teensy 4.1 ambisonics modules. A general-purpose, consumer-grade laptop computer runs the DAW which hosts the plugin, and this machine is connected to a network switch, to which our Teensy-based ambisonics modules are connected in turn. An additional Teensy serves as a time authority for the network and the general purpose machine. Figure 1 illustrates an example third order HOA installation based on our system; this system is described in further detail below.

3.1 The Time Authority

Fundamental to our system is the notion of an *authoritative source of time*. To this end, we use one Teensy as a dual-purpose USB audio device and PTP clock authority. In serving this dual purpose, the time authority ensures that the computer to which it is connected via USB, and the network of MCU-based ambisonics modules, for which it serves as the PTP authority, remain in sync. The computer, time authority, and ambisonics modules are all connected to a MikroTik CRS326-24G-2S+IN network switch, [2] which acts as a PTP boundary clock, i.e. the switch synchronises to the time authority and the ambisonics modules synchronise to the switch.

3.2 The Audio Plugin

The plugin was created using the CLAP [3] format and is hosted by a DAW running on a general-purpose laptop computer. Due to its support for the CLAP format (which provides greater flexibility with regard to channel layout than VST3), our preferred DAW has been Reaper, [4] running on either Linux or Mac OS.

Present in the plugin interface are elements for selecting modules, each of which is responsible for producing output signals for two loudspeakers. Modules are selectable via their IP addresses; once assigned a position, the plugin sends the appropriate module a user datagram protocol (UDP) packet containing its decoding coefficients, i.e. the i^{th} and j^{th} rows of the decoding matrix $\mathbf{D}$ (see equation (4)).

The plugin sends the ACN-format encoded ambisonics scene over UDP multicast on address `224.4.224.4` — taken from the second *ad-hoc* block of the IPv4 Multicast Address Assignment Guidelines [20] from the Internet Assigned Numbers Authority (IANA) — and port 49152 — the first port in IANA's dynamic port range [21]. Packets contain 16 frames of 16-bit ACN-format and contain a header with a 64-bit timestamp, for synchronous reproduction, plus 8-bit channel and 16-bit frame count. [5] The packet header is thus eleven bytes in size; 16 frames of 16 channels of 16-bit ACN-format brings the overall packet size to 523 bytes, well within the standard 1500-byte UDP *Maximum Transmissible Unit* (MTU).

3.3 Ambisonics Modules

Connected to the switch are our eight MCU-based ambisonics modules. Modules are synchronised according to the

[2] `https://mikrotik.com/product/crs326_24g_2s_in`
[3] *Clever Audio Plug-in*, `https://cleveraudio.org/`
[4] `https://reaper.fm`
[5] We use the term *frame* to refer to one sample's worth of N audio channels.

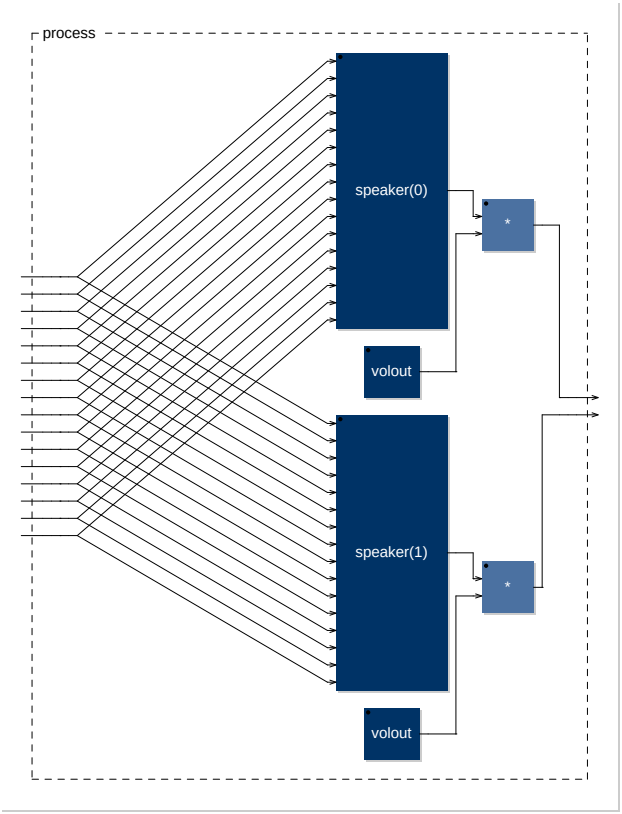


Figure 2. Faust block diagram of the distributed HOA decoder running on one of the hardware modules.

method described in our previous work [10], with a further measure in place to ensure sub-microsecond drift: in addition to using PTP-derived media clock recovery, modules check the offset of their audio subsystem from PTP time once per second and make a commensurate sampling rate adjustment. Using this hybrid adaptive clock-correction method, presentation remains synchronous amongst the network of modules to within around 200 ns at all times. The generated soundfield has been subjected to preliminary tests to ascertain its plausibility; however, no qualitative measurements or empirical assessments have been conducted at the time of writing.

An adapted version of the *AmbiTools* Faust algorithm is used to program the modules, taking advantage of the aforementioned parallelisation of the ambisonics decoder. See figure 2 for a diagrammatic representation of the distributed HOA decoder.

4. CONCLUSIONS AND OUTLOOK

In this paper we have demonstrated, taking advantage of the parallelisable nature of the decoding step of ambisonic reproduction, a distributed ambisonics system based on a low-cost MCU-based embedded computing platform. Our aim is that this should represent an important step toward a more accessible future for ambisonics-based immersive sound installations.

Though functional, scalable and modular, our system requires a comparatively laborious configuration step, that of assigning modules to their positions in the speaker array; ideally an automatic calibration mode, such as demonstrated by Wesch et al. [2] should be implemented to render this less of a pain-point. Moreover, the system presented uses the Teensy's standard audio add-on, which provides two channels of output; via time-division multiplexing (TDM), it should be possible to produce as many as 32 output channels per module, significantly simplifying the overall design providing a more favourable cost-per-output-channel.

In principle, without violating the MTU, we could support fifth order ambisonics, transmitting 36 ACN-format channels for a packet size of 1163 bytes. For optimal rather than bare-minimum reproduction, however, needing something on the order of fifty loudspeakers for a hemispherical configuration, at two speakers per module, and given the 24 ethernet ports of the CRS326-24G-2S+IN switch, this would entail daisy-chaining switches; since the switch (chosen for its low cost, high port-count, and PTP-compliance), is only capable of acting as a boundary clock, modules connected to the second switch would be subject to a delay of perhaps a few hundred nanoseconds. Again, channels-per-module improvements via TDM would sidestep this issue. Nonetheless, that, and other technical and logistical issues remain to be explored in future work.

The work presented here forms part of a broader project engaged with frugal, distributed approaches to spatial audio and virtual acoustics. Full details of the hardware, supporting software, and network protocol, including synchronisation, device discovery, and audio and control data transmission, will be made public and freely available as the project develops further.

5. REFERENCES

[1] R. Michon, J. Bizien, M. Popoff, and T. Risset, "Making Frugal Spatial Audio Systems Using Field-Programmable Gate Arrays," in *Proceedings of the 2023 New Interfaces for Musical Expression Conference (NIME-23)*, May 2023. [Online]. Available: https://inria.hal.science/hal-04169228

[2] B. Wesch, M. Mitterhuber, and M. Kaltenbrunner, "LOTTO (Localize OTTOsonics): Automated Spatial Speaker Localization Using Tetrahedral Microphone Array," in *Proceedings of the 22nd Sound and Music Computing Conference (SMC25)*, Graz, Austria, 2025.

[3] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, May 1993. [Online]. Available: https://asa.scitation.org/doi/abs/10.1121/1.405852

[4] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, ser. Springer Topics in Signal Processing. Cham: Springer International Publishing, 2019, vol. 19.

[5] S. Devonport and R. Foss, "The Distribution of Ambisonic and Point Source Rendering to Ethernet AVB Speakers," in *Proceedings of ICSA 2019*, Ilmenau, Germany, Sep. 2019.

[6] J. A. Belloch, J. M. Badía, D. F. Larios, E. Personal, M. Ferrer, L. Fuster, M. Lupoiu, A. Gonzalez, C. León, A. M. Vidal, and E. S. Quintana-Ortí, "On the performance of a GPU-based SoC in a distributed spatial audio system," *The Journal of Supercomputing*, vol. 77, no. 7, pp. 6920–6935, Apr. 2021.

[7] T. A. Rushton, R. Michon, and S. Letz, "A Microcontroller-based Network Client Towards Distributed Spatial Audio," in *Proceedings of the Sound and Music Computing Conference (SMC-23)*, Stockholm, Sweden, 2023.

[8] T. A. Rushton, R. Michon, S. Serafin, T. Risset, and S. Letz, "Networked microcontrollers for accessible, distributed spatial audio," *Frontiers in Virtual Reality*, vol. 5, Nov. 2024.

[9] B. Quiédeville, R. Michon, T. Risset, and S. Letz, "The Space Bar: An Embedded WFS Sound System," in *JIMLAC 2025 - Journées de l'Informatique Musicale*, Lyon, France, 2025.

[10] T. Rushton, R. Michon, and T. Risset, "All Together Now: A Synchronous Platform for Distributed Spatial Audio," in *Proceedings of the 22nd Sound and Music Computing Conference (SMC25)*. Graz, Austria: Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Jul. 2025.

[11] M. Popoff, R. Michon, T. Risset, P. Cochard, S. Letz, and Y. Orlarey, "Audio DSP to FPGA Compilation: The Syfala Toolchain Approach," Univ Lyon, INSA Lyon, Inria, CITI, Grame, Emeraude, Tech. Rep. RR-9507, 2023.

[12] R. Michon, Y. Orlarey, S. Letz, D. Fober, and D. Roosenburg, "Embedded Real-Time Audio Signal Processing With Faust," in *Proceedings of the 2nd International Faust Conference (IFC-20)*, Paris, France, Dec. 2020.

[13] F. Grani, D. Di Carlo, J. Madrid Portillo, M. Girardi, R. Paisa, J. S. Banas, I. Vogiatzoglou, D. Overholt, and S. Serafin, "Gestural Control Of Wavefield synthesis," in *Proceedings of the Sound and Music Computing Conference (SMC-16)*, Hamburg, Germany, Aug. 2016.

[14] A. Carôt, H. Mahmood, and C. Hoene, "GNSS-based Sound Card Synchronization," in *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems*. IEEE, Sep. 2019, pp. 309–312.

[15] J. Schleusner, C. Fahnemann, R. Pfleiderer, and H. Blume, "Sub-Microsecond Time Synchronization for Network-Connected Microcontrollers," in *2024 IEEE International Conference on Consumer Electronics (ICCE)*. Las Vegas, NV, USA: IEEE, Jan. 2024, pp. 1–6.

[16] NXP Semiconductors, "I.MX RT1060 Processor Reference Manual Rev 3," Jul. 2021.

[17] P. Lecomte, "Ambitools: Tools For Sound Field Synthesis Using Higher Order Ambisonics - V1.0," in *Proceedings of the 1st International Faust Conference (IFC-18)*, Mainz, Germany, 2018.

[18] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone," in *Proceedings of the 120th Convention of the Audio Engineering Society*, Paris, France, May 2006.

[19] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "ambiX - A Suggested Ambisonics Format," in *Proceedings of the Ambisonics Symposium 2011*, Lexington, KY, USA, 2011.

[20] D. Meyer, M. Cotton, and L. Vegoda, "IANA Guidelines for IPv4 Multicast Address Assignments," Internet Engineering Task Force, Request for Comments RFC 5771, Mar. 2010.

[21] M. Cotton, L. Eggert, J. D. Touch, M. Westerlund, and S. Cheshire, "Internet Assigned Numbers Authority (IANA) Procedures for the Management of the Service Name and Transport Protocol Port Number Registry," Internet Engineering Task Force, Request for Comments RFC 6335, Aug. 2011.