

An Interactive Virtual Icosahedral Loudspeaker Array

Markus Zaunschirm, Matthias Frank, Franz Zotter

Institute of Electronic Music and Acoustics, University of Music and Performing Arts, 8010 Graz, Austria.

Email: zaunschirm@iem.at

Introduction

Loudspeaker systems with controllable directivity are new tools to influence the presentation of sound in a room. By individually controlling the strengths of acoustic propagation paths (direct and reflected sound) such loudspeakers allow to create auditory objects. As one example of such a device, our compact icosahedral loudspeaker system (ICO) has been employed in various concerts at different venues. At each venue, the electroacoustic compositions have been adjusted to cause similar auditory objects in the specific acoustic environment. For these artistically important adjustments, despite typically limited, a reasonable amount of time is required in which the venue and the ICO are available.

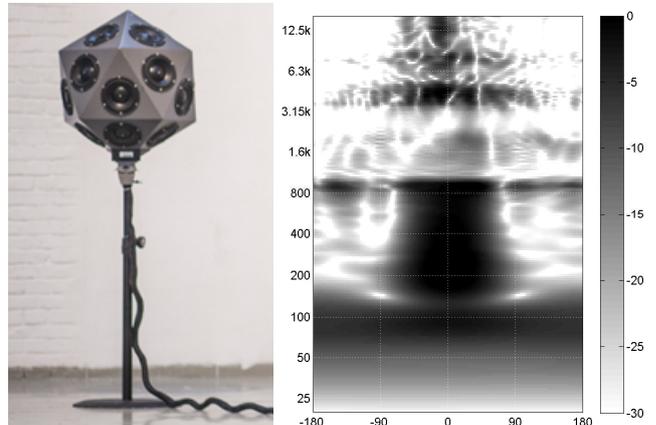
We discuss the interactive virtualization of the ICO for different playback/recording positions at various venues, to be used on a simple laptop equipped with a spatial audio workstation (Reaper with ambiX plugin suite) and headphones. The virtual ICO (VICO) employs multiple input multiple output (MIMO) convolutions comprising measured room impulse responses (RIRs), impulse responses (IRs) controlling the ICO and the higher-order microphone array (*Eigenmike EM32*), and individual head-related IRs (HRIRs) used for dynamic binaural *Ambisonic* rendering. We finally present an evaluation study that consists of listening experiments comparing the ICO and its virtualized counterpart.

The ICO

The ICO, a compact spherical loudspeaker array, houses 20 loudspeakers that are mounted into the rigid facets of an icosahedron, see fig. 1(a). Spherical beamforming of the ICO allows to synthesize well defined directivities that can be adjusted uniformly to any desired direction defined by azimuth and zenith angle. Starting from controlling the sound particle velocity on the surface of the ICO in terms of spherical harmonics, a sound pressure pattern at any radius can be derived. Accordingly, the pattern undergoes a radius- and frequency-dependent transition as it propagates. The transition of the beam pattern on the surface of the ICO to its far field counterpart is described by superposition of frequency responses for each spherical harmonic order n

$$b_n(kR) = \frac{\rho c}{i} \frac{i^{n+1}}{k h_n^{(2)}(kR)}, \quad (1)$$

where ρc is the density of air 1.2 kg/m^3 times speed of sound 343 m/s , $i = \sqrt{-1}$ is the imaginary unit, $k = 2\pi f/c$ is the wave number with the frequency f , and $h_n^{(2)}(kR)$



(a) Staging of the ICO. (b) Far-field beam pattern.

Figure 1: Staging of the ICO (left) and horizontal cut through far-field beampattern with magnitude in dB (grayscale) over polar angle and frequency for spherical beamforming with the ICO using radiation control and acoustic cross-talk cancellation (right).

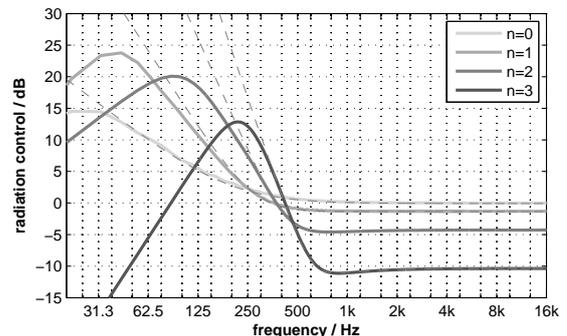


Figure 2: Analytic (dotted) and practical (solid) radial filters for the ICO. Cut-off frequencies for limiting the higher order signal amplitudes are set to $[20, 37, 115, 240]$ Hz.

is the derivative of the n^{th} -order spherical *Hankel* function of the second kind, with $R = 28.5 \text{ cm}$ as acoustically effective radius of the ICO. The patterns are strongly attenuated as the order n increases and the frequency decreases. In order to compensate for the occurrent magnitude and phase changes an equalization by *far-field radiation control* is needed, cf. fig. 2.

Moreover, the loudspeakers of the ICO share a common enclosure and thus, their motions are acoustically coupled. As directivity pattern synthesis requires individual control of these motions, a *MIMO cross-talk canceller* is required in order to compose far-field patterns out of superimposed spherical harmonics.

With the suitable control system (*far-field radiation control* and *MIMO cross-talk canceller*) the creation of desired directivities in terms of spherical harmonics is achieved by using the same tools as for arranging sounds in higher-

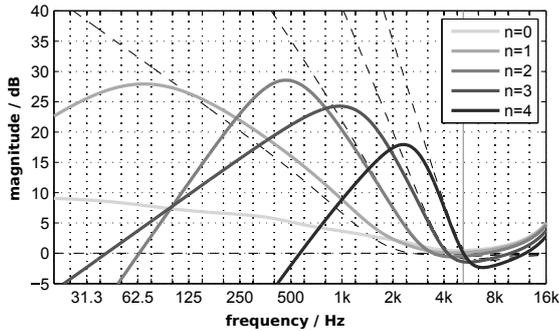


Figure 4: Analytic (dotted) and practical (solid) holographic filters for a spherical rigid sphere microphone array up to spherical harmonics order $N = 4$. Lowest possible cut-off frequencies that limit the WNG to a maximum of 20dB are [20, 59, 612, 1534, 2742]Hz.

order *Ambisonic*, i.e. *Ambisonic encoding*.

Fig. 1(b) shows the far-field beam pattern steered towards an angle of 0° on the horizon that is obtained using the above described processing. At lower frequencies the order of the directivity pattern is limited as inversion of b_n for $n > 1$ would require such a strong bass boost that would damage the loudspeakers in practice. For frequencies above 150 Hz we get a 3^{rd} -order beam pattern until the spatial aliasing becomes dominant above 800 Hz.

The virtual ICO

The signal processing chain of the interactive virtual ICO is depicted in fig. 3. A spatial arrangement of sound objects x_S , according to the desired directions φ_S, ϑ_S , is achieved by 3^{rd} -order *Ambisonic encoding*. The driving signals of the ICO's loudspeakers are then obtained by a frequency-independent decoder and an equalization using radial and crosstalk cancellation filters [1]. Until this point the processing for the ICO and its virtualized version is identical. For virtualization, the ICO to EM32 block contains a dataset of virtualized/measured MIMO room impulse responses (RIRs) of a performance situation (with an EM32 positioned at a desired listening position). Additional information on measurement configuration and data can be found here [2].

Holographic Filters

To allow for *Ambisonic* surround playback, the 32-channels of the Eigenmike EM32 are first transformed into the spherical-harmonics domain by a frequency-independent *encoder* matrix. However, the obtained signals are not suitable for direct playback as the low frequency components of the signals are mainly mapped to a omni-directional pattern (similar to the attenuation of higher-order components for the ICO) on the surface of the microphone. To compensate for the attenuation holographic filters that yield the desired *Ambisonic* playback signals are needed [3]. The frequency responses of the employed holographic filters are depicted in fig. 4, for further reading see [4].

The decomposed *Ambisonic* signals are rotated dynamically according to head movements of the listener and are finally decoded using customizable HRIRs before playback via headphones (*Ambisonic* signals of Eigenmike can also

be decoded to any loudspeaker setup using [5]).

Listening experiments

Previous work shows how sound objects generated by the ICO are perceived by listeners [6]. In this experiment we evaluated the differences of the sound objects created by the real ICO and its virtualized counterpart. The experiments are conducted in a real room with dimensions of $6.8\text{ m} \times 7.6\text{ m} \times 3\text{ m}$ and a mean reverberation time of 0.57 sec. (IEM lecture room). The ICO was placed near the front right corner of the room and the listening position is chosen approximately 4 m away from the ICO, see fig. 7.

Room Database

The 20×32 RIRs between the ICO and the Eigenmike are measured using the exponentially-swept sine technique [7] with a 5 sec. long sweep. Despite using relatively long sweeps we experienced unrealistically long reverberation times in the measured RIRs, especially for frequencies > 4 kHz. This can be explained by the limited efficiency of the ICO transducers in that frequency range. Thus, time-frequency filtering (de-noising) of the measured RIRs was needed.

Let us assume that the short-time fourier transform (STFT) of the late part of a RIR is well modeled by

$$H(t, \omega) = wv^{-t}N_1(t, \omega) + wN_2(t, \omega), \quad (2)$$

where $N_{1,2}(t, \omega)$ are two uncorrelated normalized noise processes, $v > 1$, and (t, ω) represents time and frequency dependency of the STFT, respectively. The expected value of the summed squared amplitudes (for all 640 paths) is derived from eq. 2 as

$$E\{|H(t, \omega)|^2\} = u^2v^{-2t} + w^2, \quad (3)$$

and the time where both processes are equally loud is found by

$$t_0 = \frac{\ln \frac{w}{u}}{\ln v}. \quad (4)$$

Unnatural and thus unwanted background noise present in the RIRs is adapted to the slope of reverberation by using

$$H_{denoise}(t, \omega) = \frac{H(t, \omega)}{\sqrt{1 + v^{2(t-t_0)}}}, \quad (5)$$

where the model parameters u, v, w are found at each frequency ω from comparing the modeled energy decay relief (EDR) [8] with that of the measured RIR

$$EDR_{RIR}(t) = \int_T^t |H(t, \omega)|^2 dt, \quad (6)$$

$$EDR_{model}(t) = \frac{u^2v^{-2t}}{-2 \ln v} + w^2(T - t). \quad (7)$$

The model parameters defined in eq. 2 can be found in suitable sections of EDR_{RIR} , see vertical black lines in fig. 5. Typically, the parameters u and v can be estimated by linear regression of the early part of $10 \lg EDR_{RIR}$ with

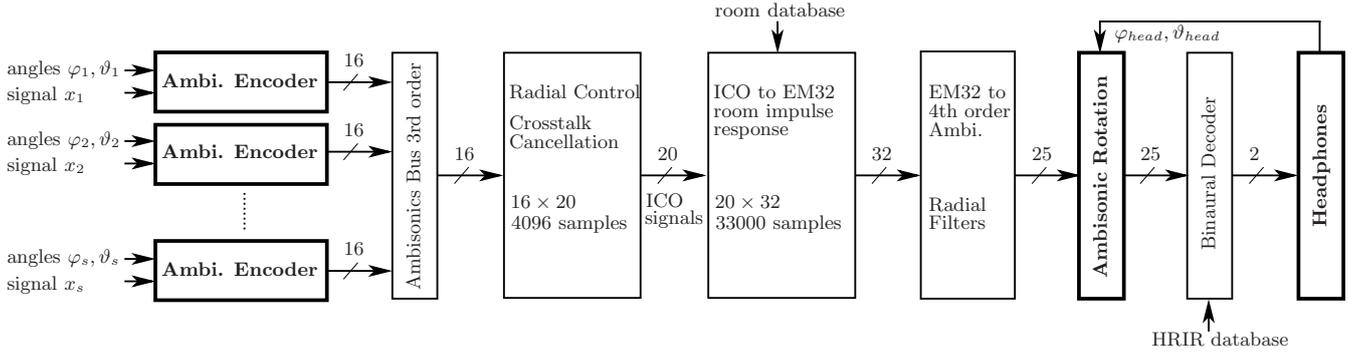


Figure 3: Processing chain of the virtual ICO containing real-time encoding and head-tracked headphone playback for user interaction (bold blocks). Sounds are spatially arranged according to the desired directions using the *Ambi. Encoder*.

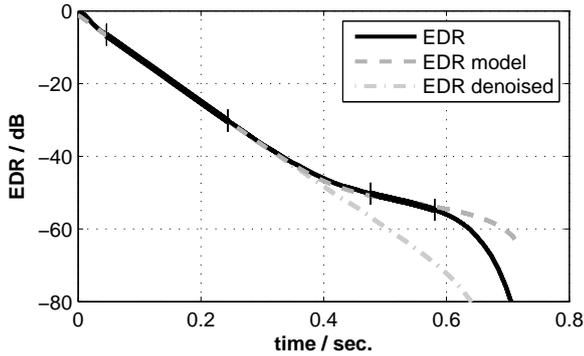


Figure 5: *EDR* of the actual, modeled and de-noised RIR for a third-octave band with center frequency $\omega = 1024$ Hz. The section for model parameter estimation is marked by vertical black lines.

$10 \lg EDR_{model} = -20t \lg v + 10 \lg \frac{u^2}{-2 \ln v}$. The parameter w can be found by regression of the later part of EDR_{RIR} by assuming $EDR_{model} = w^2(T - t)$ in that part. Figure 5 shows the *EDR* of the original, modeled and de-noised RIR for a third-octave band with a center frequency $\omega = 1024$ Hz.

Stimuli and Experiment Setup

The test signals are 1.5 sec. long and include a pink noise burst with attack and release times of $t_a = t_r = 500$ ms and a sequence of irregular short bursts. Both sounds are steered on the horizon towards 0° , 70° , 180° and -90° using a 3^{rd} order directivity pattern. The play-back signals are rendered in real time using *Reaper*, the *ambiX* and *mcfx* plugins [9] for both, the ICO and the VICO system, cf. fig. 3. Listeners are asked to specify the location of the perceived auditory object for each of the 8 (2 sounds and 4 beam directions) conditions by placing a marker corresponding to the played back sound using a GUI that includes a map of the listening room. Listeners could switch between all 8 conditions on one page arbitrarily and listen to or compare them as often as wanted. Presentation methods included playback with the real ICO and the VICO using head-tracked binaural rendering with *AKG K712* headphones. Independent of the playback method the listeners were seated at the same position and thus, could also see the real ICO for presentation using the VICO environment. For head tracking we used a *OptiTrack* system consisting of 6 *Flex 13* cameras¹. The

¹A suggestion of a smaller head tracking device can be found at

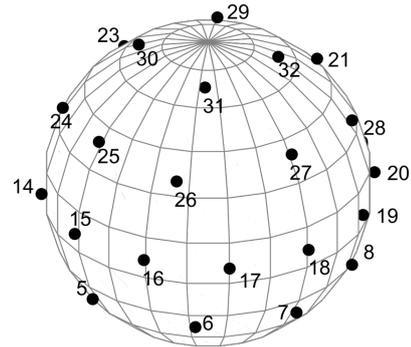


Figure 6: Virtual loudspeaker setup used for binaural rendering. The setup consists of four rings at a zenith angle of $[120^\circ, 90^\circ, 60^\circ, 25^\circ]$ including $[8, 12, 8, 4]$ loudspeakers with a spacing of $[45^\circ, 30^\circ, 45^\circ, 90^\circ]$.

Ambisonic signals were rotated contrary to head movements of listeners using an update rate of 100 Hz.

The participants were grouped such that 4 of the overall 8 participants rated first the ICO and then the VICO. The other group rated in vice versa playback order. Each playback method was repeated three times with randomly ordered conditions in each trial (3 times the 8 conditions played back with the ICO followed by 3 times the 8 conditions played back with the VICO). In each of the two groups we had two participants using individual HRIRs for playback of the VICO.

Individual in-the-ear HRIRs were measured with the blocked-ear-canal technique using a *Sennheiser KE-4-211-2* microphone and with the multiple exponential sweep method [10] in a semi-anechoic chamber. Overall, the measurement grid consisted of 1550 positions with a resolution of 2.5° and 10° in azimuth and zenith, respectively (loudspeakers were arranged on a sphere with a radius of 1.2 m). Out of the entire set of HRIRs we chose 32 that corresponded to the locations of the 32 virtual loudspeakers that are used for *Ambisonic* decoding, see fig.6. The actual frequency-independent decoder matrix is calculated according to the *All-Round Ambisonic Decoding* (AllRAD) strategy [11].

In order to evaluate not only the location of the perceived auditory object but also the naturalness of the acoustic scene for playback within the VICO environment we conducted a plausibility test. The irregular noise burst signal was steered towards the listening position and 4 partici-

pants were asked to state if they listen to the real ICO or the VICO. This time participants had to wear the headphones during the entire test. In each trial we presented the stimulus ten times in random order: five times with the ICO and VICO, respectively. The playback signal of the real ICO was adapted such that high-frequency damping of the headphone shell is accounted for.

Results

Generally, participants reported a highly natural sounding acoustic scene for playback within the VICO environment. These comments are underlined by the result of the plausibility test: only 52.5% of the ratings were correct. Figure 7 shows the 95% confidence area of the mean localization for each of the 8 conditions. It can be seen that both direction and distance are well perceivable with the virtualized version of the ICO. In detail, ratings of the perceived distance were not significantly different among all conditions. The perceived direction was only significantly different for two conditions, see bold entries in tab. 1. For both playback methods irregular noise bursts

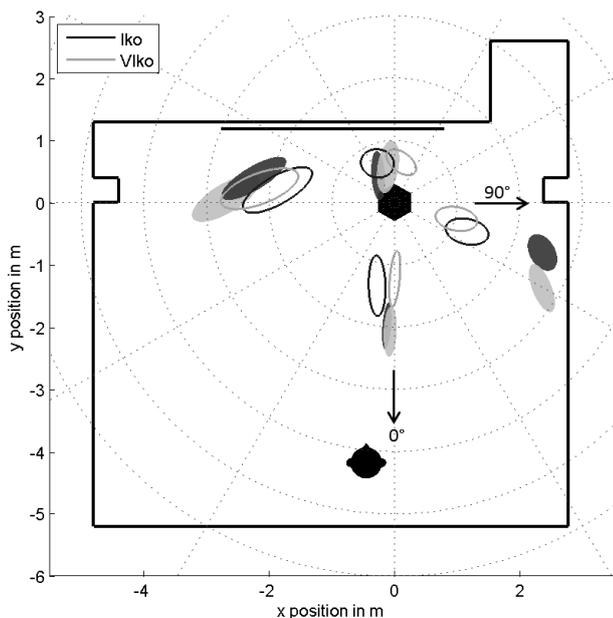


Figure 7: Setup and results of the listening experiment. The ICO (black filled hexagon) is placed in the front right corner of the room. The listening position is indicated by filled head symbol. 95% confidence ellipsoids of subjects ratings are shown for ICO (dark gray) and VICO (light gray) for a noise (filled) and irregular burst (not filled) signal.

Table 1: p -values for differences in direction and distance ratings obtained for listening experiments comparing ICO and VICO as playback device. Conditions 1 – 4 and 5 – 8 refer to static noise signal and irregular noise bursts, respectively. Within the signal groups the direction of the beam is altered clockwise starting from -90° (left).

	condition							
	1	2	3	4	5	6	7	8
direction	0.28	0.35	0.05	0.70	0.71	0.05	0.36	0.13
distance	0.24	0.24	0.22	0.71	0.62	0.81	0.32	0.37

tend to be localized nearer to the ICO/VICO than the stationary noise signal. This can be explained by a more prominent precedence effect for highly transient signals than for stationary signals [12].

Neither the use of individual HRIRs nor whether ICO or VICO was presented first had any significant impact on the ratings.

Conclusion

We presented an interactive virtualization of the ICO. Listening experiments showed that the perceived auditory objects are comparable to those obtained by the ICO. Thus, the VICO allows to evaluate new MIMO filters (for ICO and Eigenmike) and it can be used to more easily master or reproduce musical pieces for different rooms.

Acknowledgments

This work is part of the project Orchestrating Space by Icosahedral Loudspeaker (OSIL), which is funded by the Austrian Science Fund (FWF): PEEK AR 328. We thank all listeners for their participation in the experiments.

References

- [1] S. Lösler: *MIMO-Rekursivfilter für Kugelarrays*. Master thesis, 2014.
- [2] Orchestrating Space by Icosahedral Loudspeaker, URL: <http://iem.kug.ac.at/osil/>
- [3] J. Daniel and S. Moreau: Further Study of Sound Field Coding with Higher Order Ambisonics. Proc. of the 116th Convention of the Audio Eng. Soc., pp. 1–14, 2004.
- [4] S. Lösler and F. Zotter: Comprehensive Radial Filter Design for Practical higher-order Ambisonic Recording. Fortschritte der Akustik, DAGA, pp. 452–455, 2015.
- [5] F. Zotter and M. Frank: All-round ambisonic panning and decoding. AES: Journal of the Audio Engineering Society, vol. 60, no. 10, pp. 807–820, 2012.
- [6] M. Frank, G. K. Sharma, and F. Zotter: What we already know about spatialization with compact spherical arrays as variable-directivity loudspeakers. Proc. inSONIC2015, 2015.
- [7] A. Farina: Simultaneous measurement of impulse response and distortion with a swept-sine technique. Proc. AES 108th conv, no. I, pp. 1–15, 2000.
- [8] J. M. Jot: An analysis/synthesis approach to real-time artificial reverberation. Proceedings ICASSP-92, pp. 221–224 vol.2, 1992.
- [9] M. Kronlachner: Plug-in Suite for Mastering the Production and Playback in Surround Sound and Ambisonics. 136th AES Convention, April 2014, pp. 3–7, 2014.
- [10] P. Majdak, P. Balazs, and B. Laback: Multiple exponential sweep method for fast measurement of head-related transfer functions. Journal of the Audio Engineering Society, vol. 55, no. 7-8, pp. 623–636, 2007.
- [11] F. Zotter: *Holofonie für Musikinstrumente*. Fortschritte der Akustik - DAGA, 2012.
- [12] B. Rakerd and W. M. Hartmann: Localization of sound in rooms, II: The effects of a single reflecting surface. The Journal of the Acoustical Society of America, vol. 78, pp. 524–533, 1985.