Auralization of Sources with Controllable Directivity using the upscaled Source-and-Receiver-Directional Room Impulse Response

Toningenieurprojekt

B.Sc. Franck Zagala

Supervision: M.Sc. Markus Zaunschirm

Graz, September 6, 2018

kunst uni graz

institut für elektronische musik und akustik

Abstract

The assumption of spatial and temporal sparsity of acoustical reflections led to the *Spatial Decomposition Method* (SDM) for Room Impulse Responses, which was proposed in 2013 by Sakari Tervo [TPKL13]. Based on efficient B-format measurements, the SDM method allows for auralization of omnidirectional sources with a sharpened directional RIR (DRIR) at the receiver. A generalization of the SDM is proposed for the Source and Receiver Directional (SRD) RIR, which allows for auralization of sources with controllable directivity [ZBS+17]. In this project the perceptual aspects of the auralization with an efficiently measured SRD-RIR are discussed based on a comparative listening experiment involving the variable-directivity icosahedral loudspeaker (IKO).

Contents

1 Introduction

Since many years, spherical multipole expansion serves as a powerful tool to analyze, modify and synthesize three dimensional sound fields, e.g. to record or restitute an incoming wave field using *High Order Ambisonics* (HOA), analyze the directivity of a source, or to auralize an arbitrary signal in a reverberant environment.

In fact, since the radiation pattern of any source can be described using multipole expansion, just as any incoming wave field at a receiver; one could describe any source-room-receiver system as a *Source and Receiver Directional Room Impulse Response* (SRD-RIR). Note that when omitting the variable directivity of the source, we talk about a *Directional Room Impulse Response* (DRIR).

Both SRD-RIR and DRIR can be measured using a compact spherical microphone array on the receiver side and a compact spherical loudspeaker array or omnidirectional loudspeaker on the source side respectively. As adding transducers to a spherical array generally increases the controllable frequency range and the maximal order (and therefore improves the spatial resolution), one can easily be tempted to always use a very high order array to improve both the controllable frequency range and the spatial resolution.

While this approach may work for microphone arrays, it can be very challenging to mount a large amount of loudspeakers onto a small array and an increase in its size would deteriorate the temporal resolution. On the other hand, due to the frequency properties of the radial solution of the wave equation [1], it appears that the first order offers the most efficient bandwidth for a given dynamic compared to higher orders, where the little bandwidth extension comes at the cost of an unproportional hardware effort. Furthermore, a bandwidth extension could be obtained by increasing the dynamics using better quality transducers, which become affordable at low orders due to the reduced number of required transducers [Raf15].

Although these high order arrays may be beneficial for live restitution [WSF+17] or recording of complex scenes, the use of high order arrays may be avoided for the measurement of *Room Impulse Responses* (RIR). In fact, RIR measurement can profit from the time invariance and linearity of the system (which can be assumed for segmented measurements), as high order arrays can be simulated with few transducers through rotation of the array [PKDV13]. However, this technique can be time consuming and requires additional hardware such as a turntable.

In order to reduce the hardware effort while still obtaining high resolution room responses, Pulkki and Merimaa proposed the *Spatial Impulse Response Redering* (SIRR) approach in 2005. There the DRIR is decomposed in band- and time-dependent directionally varying RIRs with an additional diffuseness part [PM05]. A similar technique was proposed few years later by Tervo et al. in 2013 and is known as the *Spatial Decomposition Method* (SDM) [TPKL13]. Here the frequency dependency and diffuseness part are ommited. More recently Zaunschirm et al. proposed an extension to the SDM in [ZBS+17], which allows for variable source directivity as well, based on first order array measurement.

1. Radial filters need to amplify by ca. 6dB/Octave/Order below the spatial aliasing frequency.

The concept and determination of such an upscaled SRD DRIR is explained in Section 2. In Section 3, two listening experiments are presented which allow to investigate the perceptual aspects of the SRD-RIR compared to other auralization methods. The results of these experiments are given and discussed in Section 4, finally conclusions and outlooks are given in Section 5.

2 Source and Receiver Directional Room Impulse Response (SRD-RIR)

2.1 Concept of SRD-RIR

In a reverberant environment, any signal emitted from a source reaches the receiver through different propagation paths. This leads to a frequency dependent time-spreading of the emitted signal, which can be represented by the so called *Room Impulse Responses* (RIRs).

For many years, room acousticians used the RIR measured using omnidirectional sources and microphones to determine room acoustical parameters [fN09] or simply auralize an audio signal monorally. However, since sources are rarely omnidirectional and people perceive sound binaurally, it is obvious that this omni-to-omni RIR needs to be generalised by taking the three dimensional propagation of the sound into account.

For a source and receiver with arbitrary far-field directivity $g_R(\boldsymbol{\vartheta}_R)$ and $g_S(\boldsymbol{\vartheta}_S)$ respectively, the room impulse response (RIR) can be written as [ZBS$^+$17]:

$$h(g_r, g_s, t) = \int_{\boldsymbol{\vartheta}_R \in \mathbb{S}^2} \int_{\boldsymbol{\vartheta}_S \in \mathbb{S}^2} g_R(\boldsymbol{\vartheta}_R) h(\boldsymbol{\vartheta}_R, \boldsymbol{\vartheta}_S, t) g_R(\boldsymbol{\vartheta}_S) \ \mathrm{d}\boldsymbol{\vartheta}_R \, \mathrm{d}\boldsymbol{\vartheta}_S, \tag{1}$$

where $h(\boldsymbol{\vartheta}_R, t, \boldsymbol{\vartheta}_S)$ describes the SRD-RIR and $\boldsymbol{\vartheta}_R$ and $\boldsymbol{\vartheta}_S$ are are defined on the 2-sphere $\mathbb{S}^2$, such that:

$$\boldsymbol{\vartheta}_A = \begin{pmatrix} \sin(\varphi_A) \cdot \cos(\vartheta_A) \\ \cos(\varphi_A) \cdot \cos(\vartheta_A) \\ \sin(\vartheta_A) \end{pmatrix}, \quad A \in \{S, R\}.$$

For practical reasons, we represent the SRD-RIR and far field directivity patterns through a spherical Fourier expansion to best approximate these functions with a limited number of channels [ZBS$^+$17]:

$$h(\boldsymbol{\vartheta}_R, t, \boldsymbol{\vartheta}_S) = \sum_{n',m'} \sum_{n,m} Y_n^m(\boldsymbol{\vartheta}_R) h_{n,m}^{n',m'}(t) Y_{n'}^{m'}(\boldsymbol{\vartheta}_S), \tag{2}$$

$$g_A(\boldsymbol{\vartheta}) = \sum_{n=0}^{N_A} \sum_{m=-n}^{n} \gamma_{n,m}^A Y_n^m(\boldsymbol{\vartheta}), \qquad \text{where } A \in \{S, R\}, \tag{3}$$

where Y_n^m is the real valued Spherical Harmonics (SH) of order n and degree m and N_A denotes the maximum order.

2.2 Assumption of space-time-sparsity in the SRD-RIR

The main idea behind the SDM concept proposed by Tervo [TPKL13] lays in the assumption of temporal and spatial sparseness of the RIR. Applying the same constraint to the SRD-RIR results to:

$$h(\boldsymbol{\vartheta}_R, t, \boldsymbol{\vartheta}_S) = h_\circ(t) \cdot \delta(\boldsymbol{\Theta}_R(t) - \boldsymbol{\vartheta}_R) \cdot \delta(\boldsymbol{\Theta}_S(t) - \boldsymbol{\vartheta}_S), \tag{4}$$

$$h_{n,m}^{n',m'}(t) = h_\circ(t) \cdot Y_{n'}^{m'}(\boldsymbol{\Theta}_R(t)) \cdot Y_n^m(\boldsymbol{\Theta}_S(t)) \tag{5}$$

where $\delta(\cdot)$ denotes the Kronecker-delta function and $\boldsymbol{\Theta}_R(t)$ and $\boldsymbol{\Theta}_S(t)$ depict the *Direction of Arrival* (DOA) of the DRIR at time t on the receiver side and the *Direction of Departure* (DOD) on the source side, respectively. $h_\circ(t)$ represents the omni-to-omni RIR.

In short, each time instance of the RIR is attributed to a single propagation path in the room starting from the source with the angle $\boldsymbol{\Theta}_S(t)$ and reaching the receiver with the incidence angle $\boldsymbol{\Theta}_R(t)$.

Thus, with the omni-to-omni RIR $h_\circ(t)$ and the time dependent DOA $\boldsymbol{\Theta}_R(t)$ and DOD $\boldsymbol{\Theta}_S(t)$, a first order measurement [2] can be upscaled to any order, and thus, reach an higher spatial resolution.

2.3 Implementation

Firstly, the 1st order MIMO-RIR is measured using a home-made cubic loadspeaker array with 6 membranes (see Fig. 1(a)) and a Soundfield ST450 B-format microphone array (see Fig. 1(b)).

(a) Home-made cubic loudspeaker array

(b) Soundfield ST450 B-format microphone array

Figure 1 – First order loudspeaker and microphone arrays used for the measurement of SRD-RIR

From the first order arrays, we obtain the omni-omni response which is denoised in order to improve the SNR. The denoising algorithm is described in 2.3.1. Then the DOD and

2. Note that any spherical array configuration of higher order could also be used

DOA are computed for each time sample as described in Section 2.3.2. The upscaled SRD-RIR is obtained by (5) up to an order N_R and N_S. Finally, a time varying filter is applied to each of the channels to correct the influence of the upscaling process onto the spectrum. Fig. 2 depicts the different steps required for processing the SRD-RIR based on first order to first order measurement.

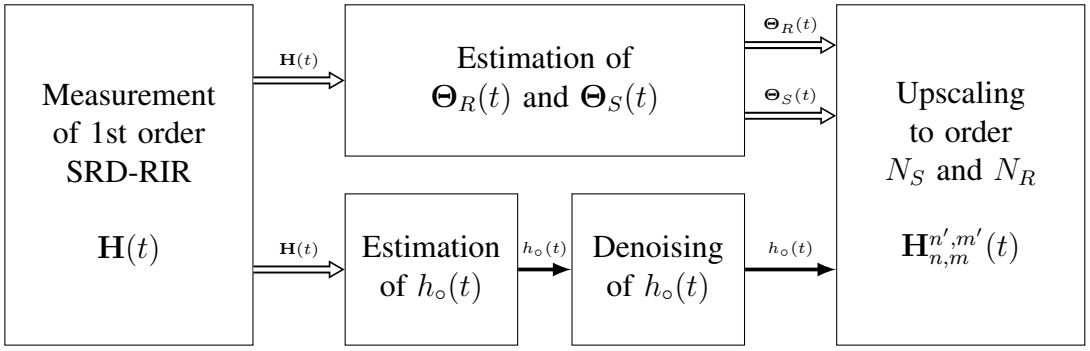


Figure 2 – Creation of a SRD-RIR of arbitrary order N_S and N_R based on a first order MIMO-DRIR

2.3.1 Denoising

The denoising of the omni-to-omni impulse response $h_\circ$ can be done as following. Assuming the impulse response can be modelled as

$$h_m(t) = an_1(t)e^{-bt} + cn_2(t),$$

where $n_1(t)$ and $n_2(t)$ are uncorrelated normed noise signals, we can numerically find the parameters a, b and c that minimize the squared error between the Energy Decay Curve $EDC_\circ$ of $h_\circ$ and the expected EDC_m of the model h_m.

$$a, b, c = \underset{a,b,c}{\arg\min} \left\| EDC_\circ - EDC_m \right\|,$$

where

$$EDC_\circ = \int_t^{T_\gg} (h_\circ(t))^2 \, \mathrm{d}t,$$

and

$$\begin{aligned} EDC_m &= \int_t^{T_\gg} \mathbb{E}\{(h_m(t))^2\} \, \mathrm{d}t \\ &= \int_t^{T_\gg} a^2 \underbrace{\mathbb{E}\{n_1^2(t)\}}_{=1} e^{-2bt} + c^2 \underbrace{\mathbb{E}\{n_2^2(t)\}}_{=1} + 2a^2c^2 \underbrace{\mathbb{E}\{n_1(t)n_2(t)\}}_{=0} e^{-bt} \\ &= \int_t^{T_\gg} a^2 e^{-2bt} + c^2 \end{aligned}$$

$$= \left[\frac{a^2 e^{-2bt}}{-2b} + c^2 t \right]_t^{T_\gg}$$
$$EDC_m = \frac{a^2 e^{-2bt}}{2b} + c^2 (T_\gg - t).$$

From there, the late linear decaying part of the impulse response can be simply attenuated by amplitude weighting. The denoised impulse response takes the form

$$h_\circ(t) \leftarrow \frac{1}{\sqrt{1 + \frac{c^2}{a^2} e^{2bt}}} \cdot h_\circ(t).$$

Here, this denoising process is applied on third-octave bands.

2.3.2 DOA and DOD estimation

Many DOA estimation algorithms are known from literature [JHN10, ZZF07, TP15] and all of them can be used for the estimation of both the DOA and DOD, thanks to the reciprocity principle. Within this work, two very simple algorithms have been implemented for the DOD and DOA estimation, respectively.

DOA estimation On the receiver side, the DOA is estimated using the *Pseudo Intensity Vector* (PIV) [JHN10].

$$\boldsymbol{\Theta}_R \propto F_T \left\{ F_b\{h_\circ(t)\} F_b \left\{ \begin{pmatrix} h_{\circ,x}(t) \\ h_{\circ,y}(t) \\ h_{\circ,z}(t) \end{pmatrix} \right\} \right\},$$

where $h_\circ$ is the omni-to-omni impulse response, $h_{\circ,x}, h_{\circ,y}, h_{\circ,z}$ are the impulse responses corresponding to the x, y and z dipoles at the receiver for an omnidirectional source, F_b depicts a zero phase band pass filter and F_T is a zero phase moving average filter, whose span covers the time propagation between the maximal distance between 2 capsules.

DOD estimation On the source side, the DOD is determined based on the *Transformed Magnitude Sensor Response* (TMSR) [ZBS+17].

$$\boldsymbol{\Theta}_S(t) \propto \sum_{l=1}^{L} F_b\{h_{l,\circ}(t)\}^2 \boldsymbol{\theta}_l,$$

where $\boldsymbol{\theta}_l$ is the position and L is the total number of loudspeakers, $h_{l,\circ}$ the impulse response of the the l-th loudspeaker measured by an omni microphone.

Both the DOA and the DOD are calculated from the measured RIRs in a frequency range between 100 Hz and 3 kHz. Furthermore, to avoid possible bias due to bad microphone-calibration, especially when the RIRs become spatially diffuse, it was proposed to generate random DOAs and DODs after ca. 300 ms [ZBS+17]. Finally the computed DOA and DOD were smoothed appropriately in order to avoid irregularities.

2.3.3 Spectral correction

The amplitude modulation occurring in each channel n, m and n', m' due to the time variation of $\mathbf{\Theta}_S(t)$ and $\mathbf{\Theta}_R(t)$ in (5) tends to whiten the upscaled SRD-RIR increasingly with the order n and n'. Therefore a spectral correction is necessary.

As described in [ZFZ18], a time variant correction filtering is applied by modulating the amplitude of each third-octave band pass filtered signal $F_b\{\tilde{h}_{n,m}(t)\}$ of an SDM-RIR. The amplitude modulation for each order n takes the form:

$$w_n^b(t) = \sqrt{\frac{2n+1}{4\pi}} \sqrt{\frac{F_T\{F_b\{h_\circ(t)\}^2\}}{\sum_{m=-n}^{n} F_T\{F_b\{h_{n,m}(t)\}^2\}}}, \tag{6}$$

where $F_T\{\cdot\}$ denotes the moving average filtering operation of span T and $F_b\{\cdot\}$ a third-octave band pass filtering.

This can be easily adapted to the SRD-RIR case. From (5), we can write

$$\left(h_{n,m}^{n',m'}(t)\right)^2 = |Y_{n,m}(\mathbf{\Theta}_S(t))|^2 \cdot |Y_{n',m'}(\mathbf{\Theta}_R(t))|^2 \cdot h_\circ^2(t), \tag{7}$$

it follows that

$$\sum_{m=-n}^{n} \sum_{m'=-n'}^{n'} \left(h_{n,m}^{n',m'}(t)\right)^2 = \sum_{m=-n}^{n} |Y_{n,m}(\mathbf{\Theta}_S(t))|^2 \cdot \sum_{m'=-n'}^{n'} |Y_{n',m'}(\mathbf{\Theta}_R(t))|^2 \cdot h_\circ^2(t). \tag{8}$$

By definition, we use orthonormal spherical harmonics $\int_{\boldsymbol{\vartheta}\in\mathbb{S}^2} Y_{n_1}^{m_1}(\boldsymbol{\vartheta}) Y_{n_2}^{m_2 *}(\boldsymbol{\vartheta})\, \mathrm{d}\boldsymbol{\vartheta} = \delta(n1 - n2) \cdot \delta(m1 - m2)$, therefore we can apply the Unsöld's Theorem [Uns27]

$$\sum_{m=-n}^{n} |Y_{n,m}(\boldsymbol{\vartheta})|^2 = \frac{2n+1}{4\pi}, \quad \forall \boldsymbol{\vartheta} \in \mathbb{S}^2. \tag{9}$$

Then, inserting (9) into (8) leads to

$$\sum_{m=-n}^{n} \sum_{m'=-n'}^{n'} \left(h_{n,m}^{n',m'}(t)\right)^2 = \frac{(2n+1)(2n'+1)}{16\pi^2} h_\circ^2(t) \tag{10}$$

This property stays unaffected when summed over time, thus we write it for a time frame of length $T+1$ centered on τ,

$$\sum_{t=-T/2}^{T/2} \sum_{m=-n}^{n} \sum_{m'=-n'}^{n'} \left(h_{n,m}^{n',m'}(t+\tau)\right)^2 = \frac{(2n+1)(2n'+1)}{16\pi^2} \sum_{t=-T/2}^{T/2} h_\circ^2(t+\tau) \tag{11}$$

which can be Fourier-transformed, then the Parseval theorem yields

$$\sum_{m=-n}^{n} \sum_{m'=-n'}^{n'} \sum_{k=0}^{T} \left(H_{n,m}^{n',m'}(k,\tau)\right)^2 = \frac{(2n+1)(2n'+1)}{16\pi^2} \sum_{k=0}^{T} |H_\circ(k,\tau)|^2. \tag{12}$$

One could now define a time varying equalizer $W_{n,n'}(k,\tau)$ in frequency domain, such that (12) is satisfied,

$$\tilde{H}_{n,m}^{n',m'}(k,\tau) = W_{n,n'}(k,\tau) \cdot H_{n,m}^{n',m'}(k,\tau), \tag{13}$$

therefore, by inserting (13) into (12), we obtain

$$\sum_{m=-n}^{n} \sum_{m'=-n'}^{n'} |W_{n,n',\tau} \cdot H_{n,m}^{n',m'}(k,\tau)|^2 = \frac{(2n+1)(2n'+1)}{16\pi^2} |H_\circ(k,\tau)|^2, \tag{14}$$

thus, the time varying equalizer has the following amplitude spectrum

$$|W_{n,n',\tau}| = \sqrt{\frac{(2n+1)(2n'+1)}{16\pi^2}} \cdot \sqrt{\frac{|H_\circ(k,\tau)|^2}{\sum_{m=-n}^{n} \sum_{m'=-n'}^{n'} |H_{n,m}^{n',m'}(k,\tau)|^2}}. \tag{15}$$

Practically, this can be implemented similarly to (6), by applying a time varying amplitude modulation to each band-pass-filtered signals in order to guarantee a spectrally smooth correction:

$$w_{n,n'}^{b}(t) = \sqrt{\frac{(2n+1)(2n'+1)}{16\pi^2}} \sqrt{\frac{F_T\{F_b\{h_\circ(t)\}^2\}}{\sum_{m=-n}^{n} \sum_{m'=-n'}^{n'} F_T\{F_b\{h_{n',m'}(t)\}^2\}}}, \tag{16}$$

$$\tilde{h}_{n,m}^{n',m'}(t) = \sum_{b} F_b\{h_{n,m}^{n',m'}(t)\} \cdot w_{n,n'}^{b}(t). \tag{17}$$

3 Listening experiment

In order to investigate the perceptual properties of the SRD-RIR, a preliminary listening experiment is conducted (Section 3.2) in order to investigate the potential of the SRD algorithm and implement some improvement for a second listening experiment (Section 3.3).

In both listening experiments, participants were asked to indicate the position of the perceived source location.

As physical source we used a variable directivity speaker in a reverberant environment. Here, a compact icosahedral loudspeaker array with 20 independent membranes (IKO) was used, enabling a variable directivity up to 3rd order [ZZFK17]. It was used as a beamformer with a max-r_E weighted directivity pattern. The max-r_E directivity maximizes the norm of the centroid of the squared directivity pattern along its beam direction $\boldsymbol{\theta}_S \in \mathbb{S}^2$ (see Fig. 3), which is approximated by [ZF12] [3]:

$$g_S(\boldsymbol{\vartheta}) = \sum_{n=0}^{N_S} \frac{2n+1}{4\pi} a_n P_n\left(\langle \boldsymbol{\theta}_S, \boldsymbol{\vartheta} \rangle\right), \quad \boldsymbol{\theta}_S, \boldsymbol{\vartheta} \in \mathbb{S}^2, \tag{18}$$

with

$$a_n \approx P_n\left(\cos\left(\frac{137.9^\circ}{N_S + 1.51}\right)\right), \tag{19}$$

3. Note that (18) is a special case of (3)

where, $P_n(\cdot)$ denotes the Legendre Polynomial or order n and $\langle \cdot, \cdot \rangle$ is the scalar product operator.

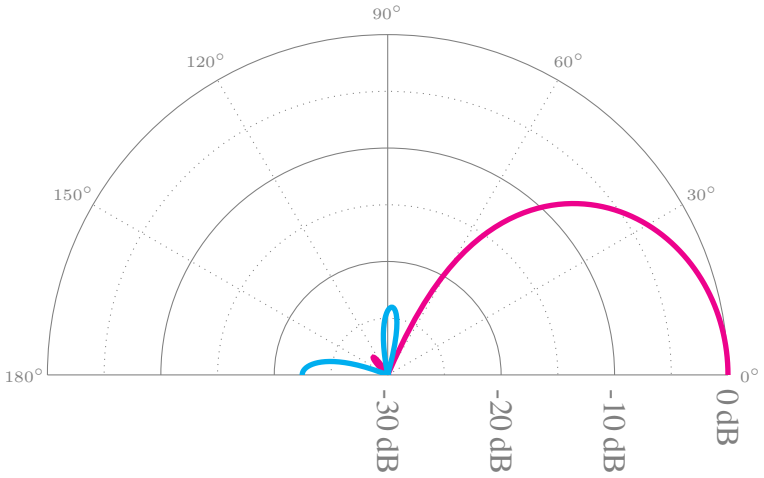


Figure 3 – max-r_E directivity of order 3

As the beam direction lies on the horizontal plane during all experiments, from now on, the beam direction $\boldsymbol{\theta}_S$ will simply be indicated by its azimutal angle ϕ_S.

In both conducted experiments, the localisation performance of the phantom source created with the upscaled SRD algorithm as proposed in Section 2, was compared to other auralization techniques.

3.1 Auralization techniques

For comparison, different auralization techniques have been used within both listening experiments.

IKO: IKO is directly auralised *in-situ*. The block-diagram is presented in Fig. 4(a).

SRD vIKO: SRD-RIR are computed for an 7th (Experiment I) and 15th (Experiment II) order input and 5th order output as described in Section 2.3. For the measurement, a home-made cubic loudspeaker array with 6 membranes was used, the receiver consists of a Soundfield ST450 MKII.

In order to virtually introduce IKO within the system, the inputs of the SRD-RIR were filtered by the directivity of IKO represented in SHs, corresponding to 7th or 15th order measurements of the IKO [SZZ18]. The scene in ambisonics is then rendered in real-time using a *state-of-the-art* renderer [ZSH18] using the *Head Related Transfer Functions* (HRTF) of a dummy head Neumann KU100 and equalised AKG 702 as headphones, while the head position was tracked with a MrHeadTracker [RBF+17]. The block-diagram is presented in Fig. 4(b).

SRD: Similarly to SRD vIKO, the same algorithm without the inclusion of a virtual IKO was also examined in experiment II. The source order was set to $N_S = 3$.

SDM vIKO: SDM-RIR [TPKL13] was measured for each membrane of IKO separately with an Soundfield ST450 MKII. The auralization over headphones occurs as for the SRD vIKO. The block-diagram is presented in Fig. 4(c).

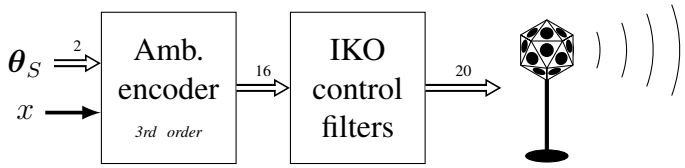
Dummy Head vIKO: IKO was measured using a dummy head Neumann KU100, this enables the most direct auralization of IKO using headphones. However, to enable a dynamic reproduction with head tracking, BRIRs were measured for different orientations,

$$\phi_{head} \in \{-45°, -30°, -15°, 0°, 15°, 30°, 45°\}.$$

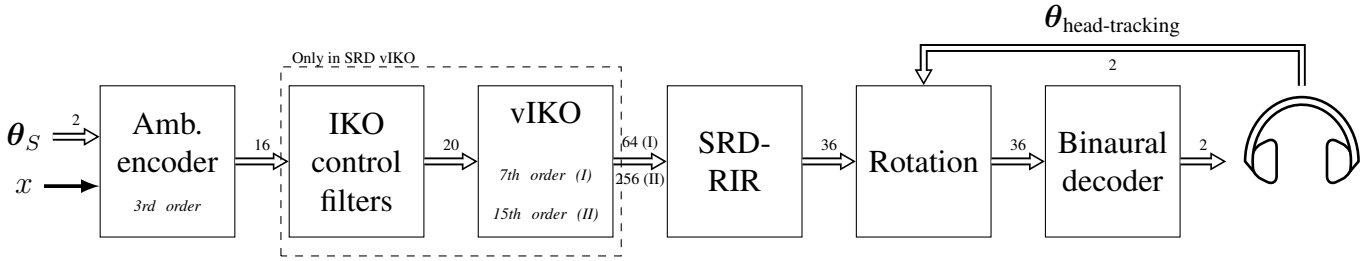
Thus, a *Motion Tracked Binaural* (MTB) decoding based on Lindau [HAD06, LR10] could be implemented. For $f < 3$kHz, the resulting BRIR for a given head orientation consists of a linear interpolation in time domain between the BRIR of the 2 closest measured orientations. For $f > 3$kHz, the linear interpolation operates on the magnitude in frequency domain, while the phase information were replaced by the phase of the closest BRIR. In order to avoid the phase to change to often when the head orientation lies between 2 measurements angles, an hysteresis was introduced. As for SRD and SDM vIKO, equalised AKG 702 headphones were used. The block-diagram is presented in Fig. 4(d).

Eigenmike vIKO: To compare against a direct higher order measurement the MIMO responses between the IKO and an Eigenmike EM32 were measured as well. The Eigenmike was encoded as proposed in [Lös13]. The Auralization via headphones is similar to SRD and SDM. The blockdiagram is presented in Fig. 4(e).

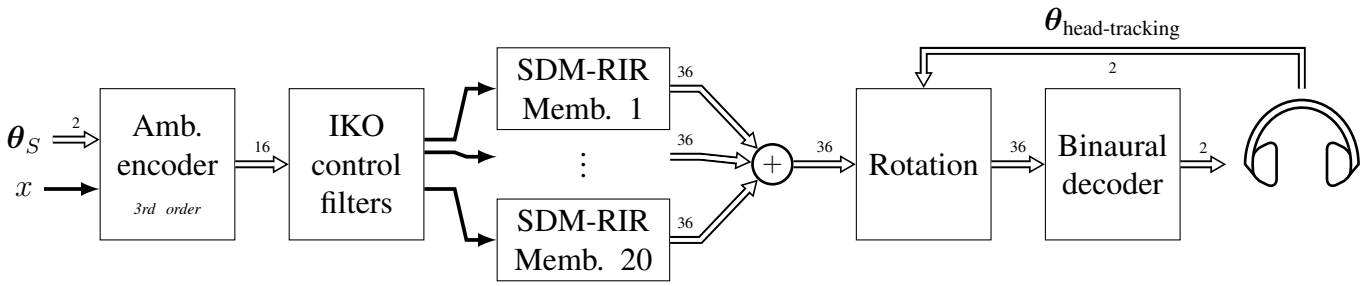
Spectral correction A spectral correction was introduced for most auralization methods (except for direct auralization of IKO) in order to avoid localisation bias induced by spectral properties. For all beam directions $\phi_S \in \{0, 10°, \ldots, 350°\}$, the spectral difference of a binaurally auralised broad band pink noise was compared to a reference one (SDM vIKO for listening experiment I, Dummyhead vIKO for listening experiment II). A minimum phase filter was designed according to the average spectral amplitude difference over all angles after third-octave smoothing.



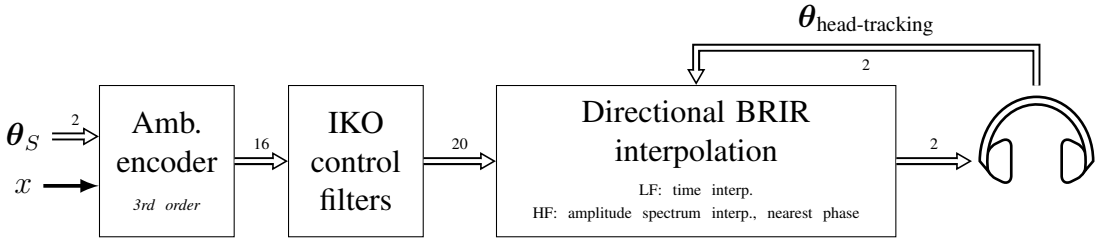
(a) IKO,



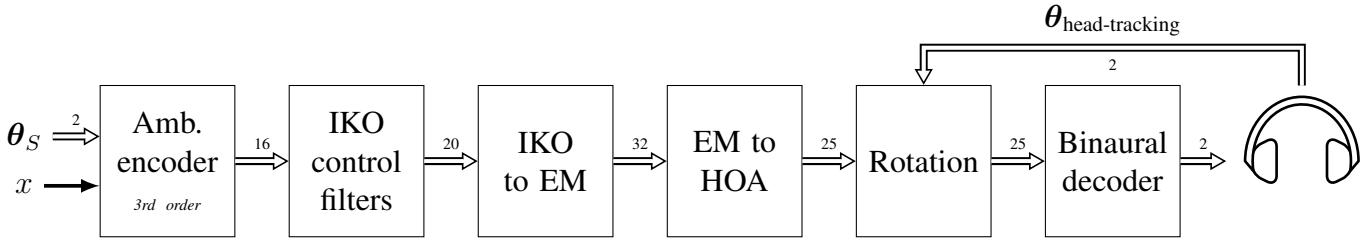
(b) SRD and SRD vIKO,



(c) SDM vIKO,



(d) dummyHead vIKO,



(e) Eigenmike vIKO,

Figure 4 – Block-diagrams of the different auralization techniques

3.2 Listening experiment I

Listening experiment I was conducted in the classroom of the IEM (the geometry as well as the coordinates system is given in Fig. 5), the IKO was at position $(-3.1, 5.36)$m, and the listener at the position $(-3.6, 2)$m.

The listening experiment was decomposed into *scenes*, where a *Graphical User Interface* (GUI) enables the participant to move pointers onto a map of the room in order to indicate the position of the perceived source.

The GUI was written in MATLAB, while the signals were processed in REAPER, using the Kronlachner plugins *ambix* and *mcfx*. The beam direction was controlled by automations and the coordination between REAPER and the GUI was realised with OSC messages.

As test signals, we used (i) pink noise and (ii) a 1.3s brass sample, auralised either with a real IKO, SRD-vIKO or SDM-vIKO and the beam direction was switched sequentially every 1.5s with a linear transition of about 0.2s.

The participants (10 researchers and master students of the IEM, all experienced with spatial hearing tests) were asked to place a pointer on a map according to the perceived localisation of the source for each of the beam directions.

In total 6 stimuli were played repetitively, where each one of them was auralised for a specific beam direction

$$\phi_S \in \{-180^\circ, -110^\circ, -60^\circ, 0^\circ, 60^\circ, 110^\circ\}.$$

The experiment starts with three training sessions (pink noise signals with all three auralization techniques), then the test consists of 3 auralization techniques $\times$ 2 signal types $\times$ 3 repetitions = 18 scenes, randomly ordered.

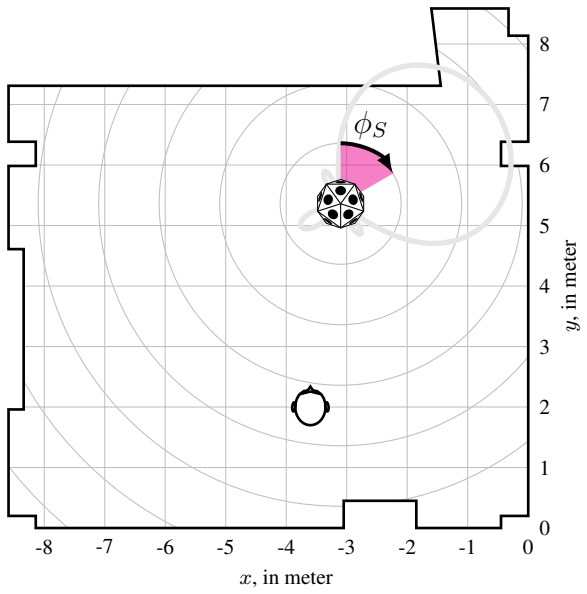


Figure 5 – Environment for the listening experiment A. $\phi_S = 60°$.

3.3 Listening experiment II

In order to suppress the possible bias occurring in the first experiment (e.g. due to the identical sequence of the beam direction and the obligation to use or remove the headphones depending of the auralization type), and try to improve the SRD vIKO restitution base on the results of experiment I, by increasing the order of the virtual IKO, a second experiment was undertaken, this time using measurement of the György-Ligeti-Saal, Graz.

Furthermore, for organisational reasons and better reproducibility, the second experiment took place in a virtual environment using the HTC vive VR glasses and AKG 702 headphones. The hall was modeled in Unity based on the buildings plans and photogrammetry for the purpose to be as realistic as possible. The head-tracker previously used in experiment I has been replaced by the built-in tracker of HTC-vive.

Similarly to the first experiment, the users are asked to indicate the perceived position for each of the 5 stimuli per scene from each scene, which consisted of pink noise bursts. An intuitive interface enables the participants to place markers at any position in the room with the help of the HTC controller by pointing towards the desired direction and adjusting the distance with a trackpad. The virtual environment was equipped with buttons on the floor, that could be actioned in order to toggle one of the five stimuli, play or pause the audio or go to the next scene.

Each scene consisted of either 5 different beam directions for a given auralization technique or 5 different auralization techniques for the same beam direction. Hereby the 5 auralization techniques consist of SRD vIKO, SRD (without IKO), SDM vIKO, Dummyhead vIKO and Eigenmike vIKO.

The 5 different beam directions are chosen in order to excite the first order reflection of a reflector and are

$$\phi_S \in \{0°, 36°, 82°, 180°, -90°\}.$$

As each scene was introduced 2 times, each of the 13 test persons (all researchers and master students of the IEM and experienced with spatial hearing tests) was confronted to 2 repetitions $\times$ 10 scenes $\times$ 5 stimuli = 100 stimuli to localise.

The equipment was setup in the middle of the room and four 0.94×1.88m reflectors were installed in order to reflect the source signal to the listener (see Fig. 6).

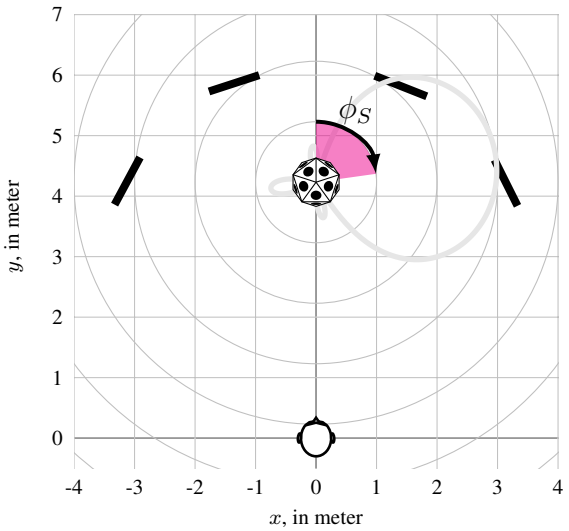


Figure 6 – Environment for the listening experiment B. $\phi_S = 82°$.

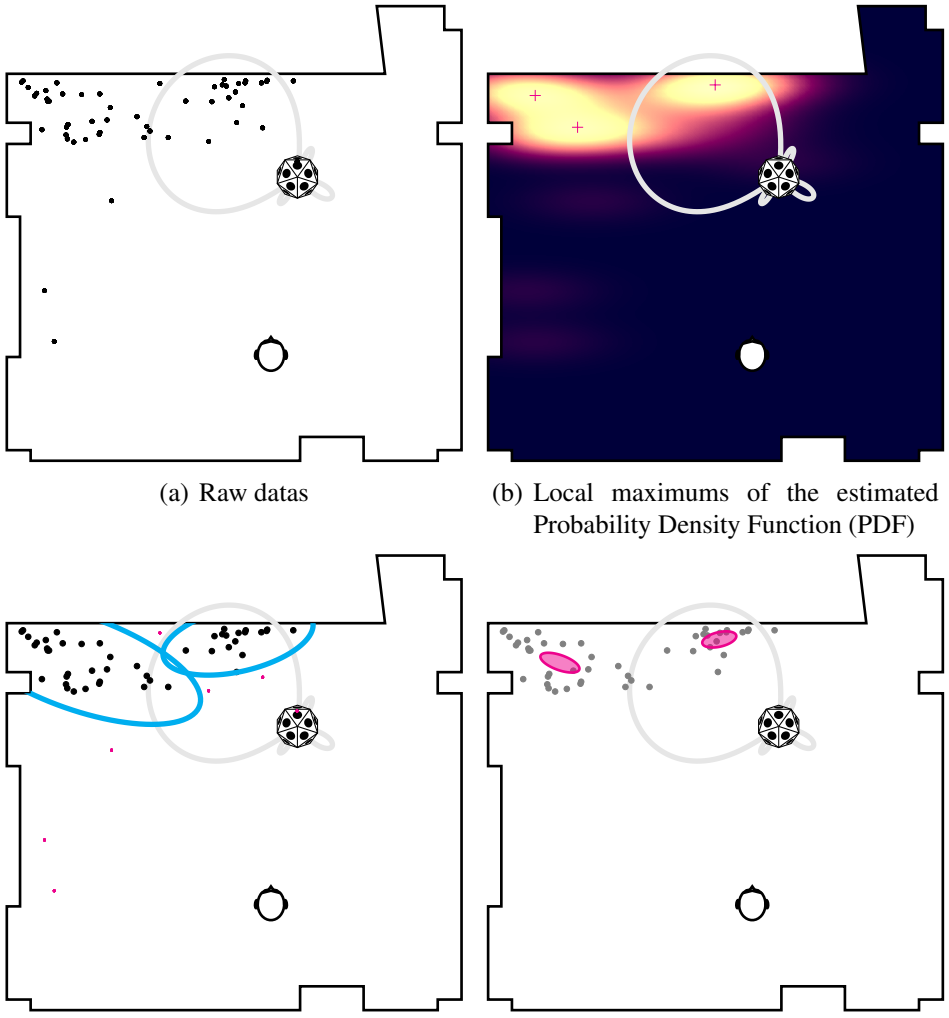
Figure 7 – 360° photograph of the acoustical environment from the listener point of view in listening experiment II, György Ligeti Saal, Graz

Figure 8 – Screen shot of the virtual environment from the listener point of view in listening experiment II

4 Results

4.1 Data processing

In order to express the 2D confidence interval of the phantom sources position, an novel algorithm is proposed. First, assuming a possible multimodal distribution of the results, the modes which corresponds to phantom sources are identified by finding the points who most probably belong to them Fig. 9(c), then the new datasets can be used to compute the confidence intervals Fig. 9(d).



(a) Raw datas

(b) Local maximums of the estimated Probability Density Function (PDF)

(c) Iterative determination of rejection ellipses for $\sigma^2 = 3$ and merging of close modes

(d) Determination of the 95% confidence ellipse for each mode

Figure 9 – Data processing algorithm in the case of multimodal distributed datas. Datas from experiment 1 using IKO , $\phi_S = -60°$.

4.1.1 Outliers removal

The detection of outliers can often be challenging, since it appears that the results may not necessarily be Gaussian distributed. In fact, in some cases, multiple distinct phantom sources are perceived, leading to multimodal distributed datas. In order to overcome those challenges, the main modes (modes identified as having at least half the value of the maximum mode in the estimated *Probability Density Function* (PDF) (see Fig. 9(b)) were analysed separately, hence each point may be either assumed as outlier, part of one, or more main modes.

For each detected mode, a set $\mathcal{X}_{\text{kept}}$ based on the 15 closest points is iteratively widened by the next closest point until it does not lie within the $\sigma < 3$-ellipse, which is actualised within each iteration. For simplicity, the remaining points in $\mathcal{X}_{\text{rem}}$ are transformed in a space where, $\mathcal{X}_{\text{kept}}$ is distributed according to a standard gaussian distribution ($\boldsymbol{\mu} = (0,\, 0)^\top$, $\boldsymbol{\Sigma} = \mathbf{I}_2$), thus the next probable point is the one whose transformation leads to the least norm.

The algorithm for the outliers detection is described in Algorithm 1.

4.1.2 Computation of 95% confidence regions

Assuming each distribution mode is Gaussian distributed, it is convenient to compute their 95%-confidence region. This can be easily done by working in a space were transformed points are normally distributed with zero-mean and unit variance. In this case the 95% confidence radius can be computed as following [JW07]:

$$r_{95\%} = \frac{p(N-1)}{N \cdot (N-p)} F_{p,N-p}(0.95) \tag{20}$$

where,

$p = 2$	space dimension
$F_{p,N-p}(\alpha)$	inverse F distribution for α-confidence and parameters p and $N-p$
N	number of points in the mode.

The 95%-confidence ellipse can then be numerically determined by transforming a circle of radius $r_{95\%}$ back in the original space:

$$\mathbf{E} = \frac{1}{\sqrt{N-1}} \mathbf{U} \mathbf{S}_{2\times 2} \mathbf{C}, \tag{21}$$

where,

$\mathbf{E}$	$[2 \times N_E]$	coordinates of the points forming the ellipse in original space
$\mathbf{C}$	$[2 \times N_E]$	coordinates of the points forming the circle in transformed space
N_E		number of points forming the ellipse
N		number of points in the mode.

input : set $\mathcal{X}_{\text{all}}$ containing all the points of the scenario
output: sets $\mathcal{X}_m$ containing the retained points for each main mode $m \in \{1, \dots M\}$

Estimate PDF from $\mathcal{X}_{\text{all}}$ (e.g. with Kernel density Estimation)
Find the modes of the estimated PDF (e.g. at least half the value of the maximum mode)
foreach *mode detected* **do**
    Create set $\mathcal{X}_{\text{kept}}$ containing the 15 closest points in $\mathcal{X}_{\text{all}}$ from the current mode
    Create set $\mathcal{X}_{\text{rem}} = \mathcal{X}_{\text{all}} - \mathcal{X}_{\text{kept}}$ containing the remaining points
    $search \leftarrow$ `true`
    $m \leftarrow 0$
    while $search$ *is* `true` **do**
        Compute mean coordinates $\bar{\mathbf{x}}_{\text{kept}}$ from $\mathcal{X}_{\text{kept}}$
        Center the points: $\mathcal{X}_{\text{mean-free}} = \{\mathbf{x} - \bar{\mathbf{x}}_{\text{kept}} | \mathbf{x} \in \mathcal{X}_{\text{kept}}\}$
        Organize the points of $\mathcal{X}_{\text{mean-free}}$ into a $2 \times |\mathcal{X}_{\text{mean-free}}|$-matrix $\mathbf{X}_{\text{mean-free}}$
        Decompose $\mathbf{X}_{\text{mean-free}}$ in singular values, such that $\mathbf{X}_{\text{mean-free}} = \mathbf{USV}$, where $\mathbf{U}, \mathbf{V}$ are unitary and $\mathbf{S}$ is diagonal
        Transform $\mathcal{X}_{\text{rem}}$ in other space: $\mathcal{Y}_{\text{rem}} = \{\sqrt{N-1}\mathbf{S}_{2\times 2}^{-1}\mathbf{U}^{\top}\mathbf{x} | \mathbf{x} \in \mathcal{X}_{\text{rem}}\}$
        Find the point $\mathbf{y}_{\text{new}}$ in $\mathcal{Y}_{\text{rem}}$ having the minimal norm
        if $\|\mathbf{y}_{new}\| \leq 3$ *(corresponding to the $\sigma^2 = 3$-ellipse)* **then**
            Find the point $\mathbf{x}_{\text{new}}$ in $\mathcal{X}_{\text{rem}}$ corresponding to $\mathbf{y}_{\text{new}}$ in the transformed space
            Insert the point $\mathbf{x}_{\text{new}}$ into $\mathcal{X}_{\text{kept}}$
            Remove $\mathbf{x}_{\text{new}}$ from $\mathcal{X}_{\text{rem}}$
        else
            if *current mode similar to a previous mode (e.g. >80% shared points)* **then**
                Find index l of similar mode
                Join both modes together: $\mathcal{X}_l \leftarrow \mathcal{X}_l + \mathcal{X}_{\text{kept}}$
            else
                $m \leftarrow m + 1$
                $\mathcal{X}_m \leftarrow \mathcal{X}_{\text{kept}}$
            end
            $search \leftarrow$ `false`
        end
    end
end

Algorithm 1 – Algorithm for outliers removal

4.2 Listening experiment I

The 95%-confidence regions of the perceived source localisation for each auralization technique, beam direction and both signal types are depicted in Fig. 10. The estimate of the PDF, using Kernel density estimation is given in Fig. 13.

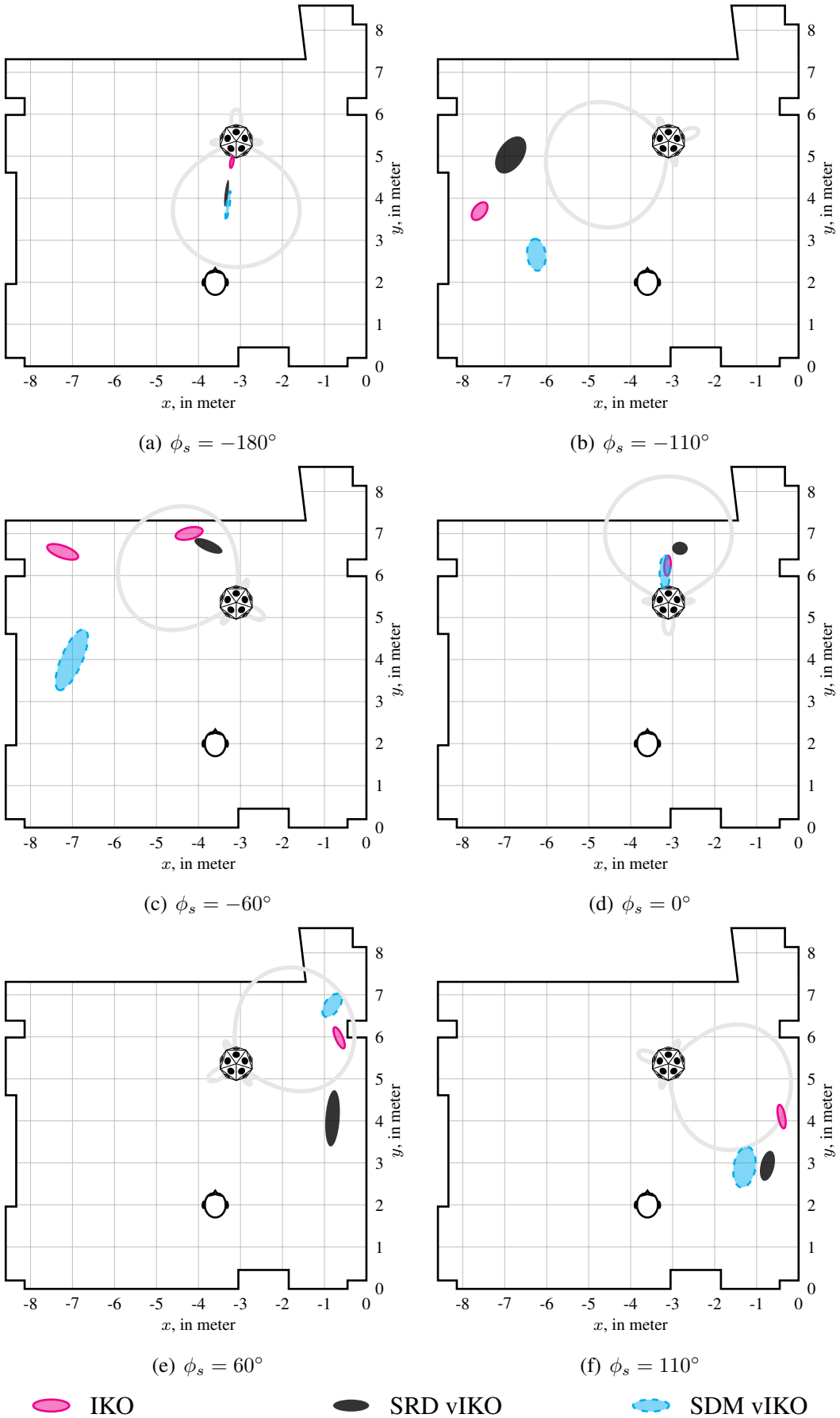
The results in Fig. 13 indicate an unimodal distribution under all test conditions, expect for $\phi_S = -60°$ and direct playback with the IKO (bimodal distribution). For that condition, listeners reported the perception of 2 distinct sources.

As desired, the stimulation of a specific wall reflection did lead to a modification of the perceived position toward the surface as expected from [ZF15, WSF$^+$17], similarly for any of the auralization techniques. However strong deviations between the different auralization techniques can be observed: given IKO as the reference, it would have been desirable for all confidence intervals to be very similar to those of IKO.

In Fig. 10(a), where the beam is pointing directly toward the listener, the perceived source direction was similar for all auralization techniques. The difference in the perceived distance may be due, on one hand, to the use of the binaural restitution for both SRD vIKO and SDM vIKO which is well known for the deterioration of the externalisation.

In the case where the beam is pointing in the opposite direction from the listener, see Fig. 10(d), the source localisations were not significantly similar in a statistical sense, but both the differences between perceived directions and distances appear reasonable.

For all other scenarios Figs. 10(b), 10(c), 10(e) and 10(f) the perceived positions do not appear reasonably similar to the reference. This strong differences may be partly explained by the use of a relatively low order vIKO (up to the the 7th order measurements) in the processing chain of SRD vIKO, which was then adjusted to 15th order in the context of the second experiment. Furthermore, the fact that the auralization of both SRD vIKO and SDM vIKO uses non-personalised HRTFs, contrary to the IKO which is perceived with the ”true ears” of the test person can also partly explain the evident differences between auralization techniques. Another source of error may also be due to the possible small differences in the placement and orientation of the source and receiver during the measurements. Lastly, the equalisation and gain adjustment may also affect the perceived localisation and especially the distance.



(a) $\phi_s = -180°$ (b) $\phi_s = -110°$

(c) $\phi_s = -60°$ (d) $\phi_s = 0°$

(e) $\phi_s = 60°$ (f) $\phi_s = 110°$

Figure 10 – 95% confidence region for each beam direction.

4.3 Listening experiment II

Fig. 11 depicts the 95%-confidence ellipses obtained during the experiment II. The estimated PDF are shown in Fig. 14.

At first sight, the Eigenmike vIKO clearly gives the most different results compared to all other auralization techniques.

Regarding the other auralization techniques, the perceived localisations appear very similar in most of the cases although the confidence interval not necessarily overlap (Figs. 11(a) to 11(c) and 11(e)).

In all scenarios, it appears that the azimuthal angle of all confidence ellipses are similar Fig. 11, except for the auralisation with Eigenmike vIKO.

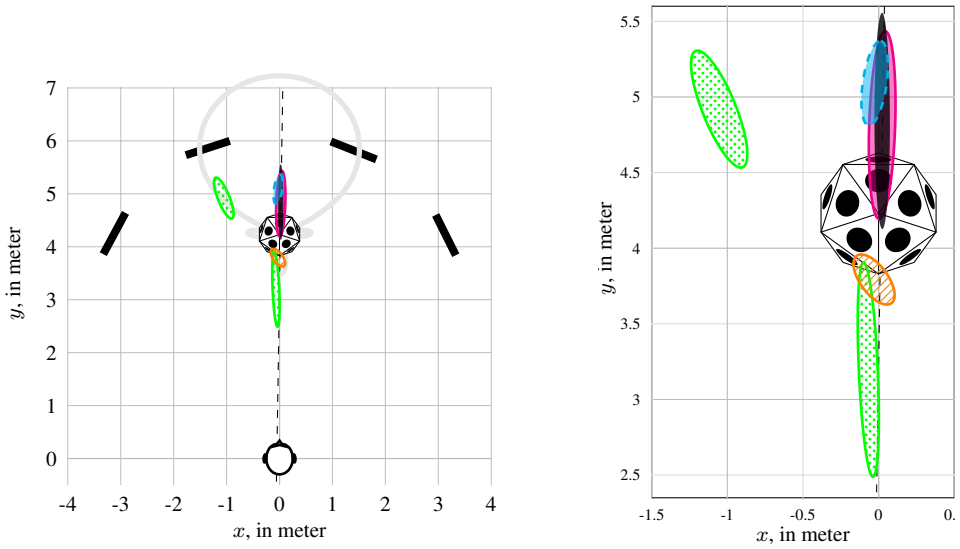
However, especially in the case where $\phi_S = \{0°, 180°\}$, the distance perception appears to spread a lot, as shown in Figs. 11(a) and 11(e).

As a first simple fidelity criterion, the average distance between the points lying within the rejection ellipse to the mean coordinate of the reference point (which is arbitrarily chosen as the mean coordinate of the points from the Dummyhead vIKO who lie within the rejection ellipse) is given in Table 1.

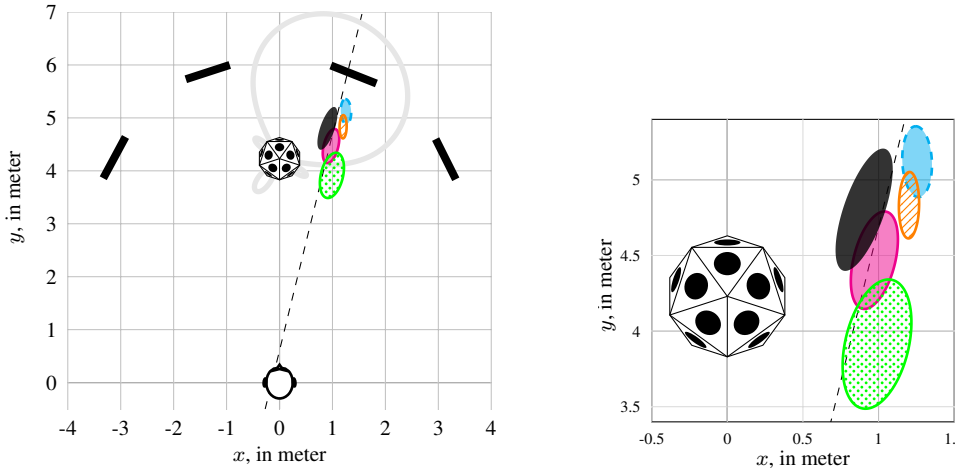
		Auralization technique				
		DummyHead vIKO	SRD	SRD vIKO	Eigenmike vIKO	SDM vIKO
ϕ_S	$0°$	1.15	1.17	0.36	1.70	1.03
	$36°$	0.74	0.97	0.76	0.01	0.56
	$82°$	0.32	0.22	0.67	1.53	0.52
	$180°$	1.10	1.33	1.47	1.55	1.18
	$-90°$	0.20	1.19	0.76	1.76	0.41

Table 1 – Mean euclidean distance between points within the rejection ellipse and the mean coordinates of reference coordinates (mean coordinates with Dummyhead vIKO), in meter.

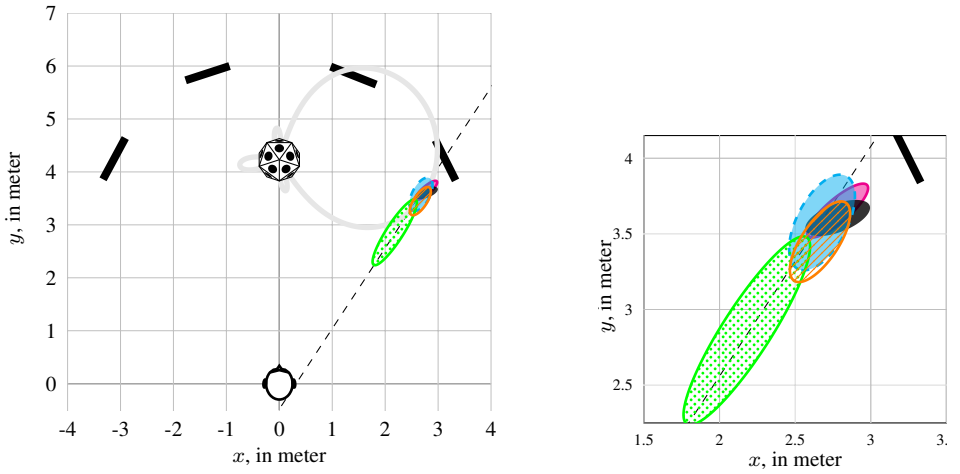
As it appears, that all results are mainly distributed along a certain axis for each beam direction (see dashed line in Fig. 11), it is proposed to use the projection coefficient of each point onto this principal axis as a simple one-dimensional fidelity criterion for each beam-direction, where the origin is set to the median coordinate obtained with Dummyhead vIKO. The principal axis is obtained by applying the singular decomposition method onto the entire data set of the given beam-direction after outlier removal. The Kernel density estimates of these projection coefficients, the 95%-confidence interval and median point are depicted in Fig. 12



(a) $\phi_s = 0°$



(b) $\phi_s = 36°$



(c) $\phi_s = 82°$

Dummy Head vIKO SRD SRD vIKO

Eigenmike vIKO SDM vIKO

Figure 11 – 95% confidence region for each beam direction.

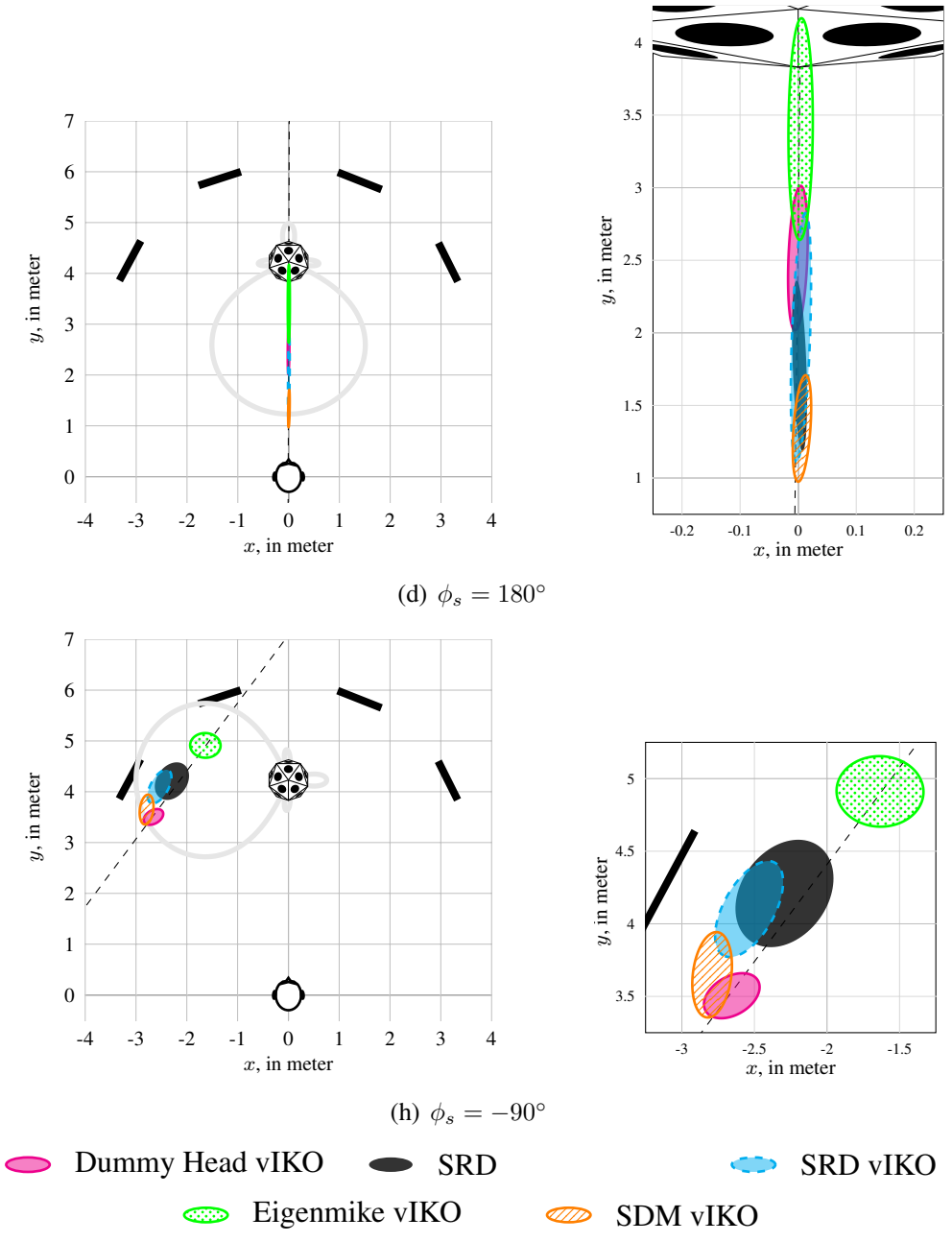
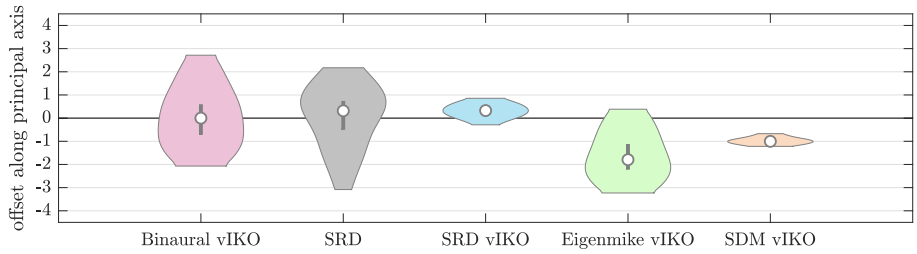
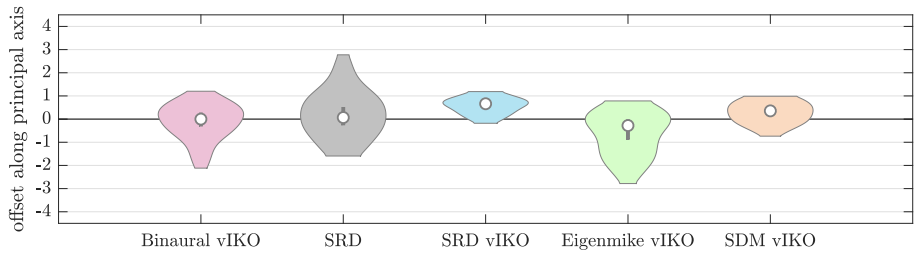


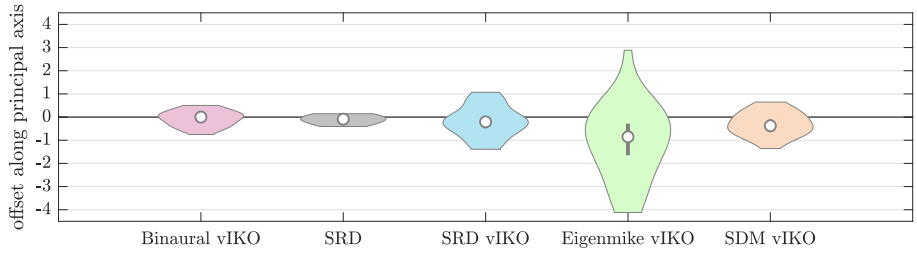
Figure 11 – 95% confidence region for each beam direction.



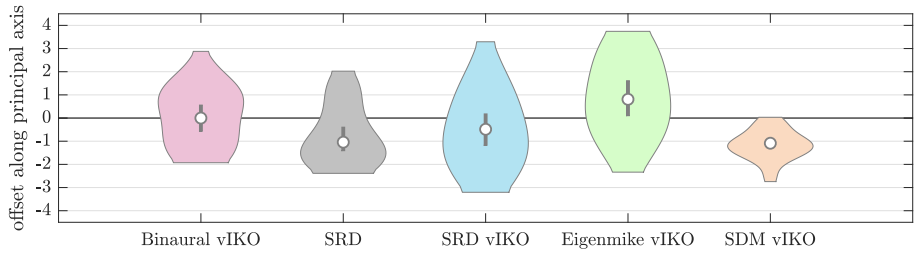
(a) $\phi_S = 0°$



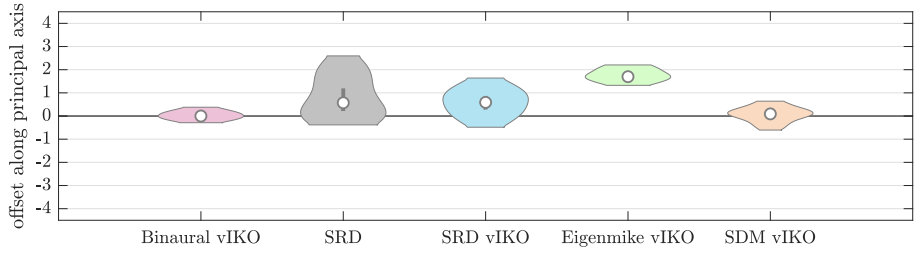
(b) $\phi_S = 36°$



(c) $\phi_S = 82°$



(d) $\phi_S = 180°$



(e) $\phi_S = -90°$

Figure 12 – Violin plot representing the kernel density distribution of the projection coefficients of each point onto the principal component axis

5 Conclusion

This work introduced the concept of an upscaled SRD-RIR and present two listening experiments (of which one preliminary test) in order to evaluate the perceived localisation of a phantom source with the SRD-RIR compared to other auralization techniques.

Although the SRD-RIR requires only first order arrays for its measurements, experiment II shows that its induced localisation of a phantom source appears to be very similar to other techniques such as SDM or head-tracked binaural recording with a Dummyhead and even outperforms high order MIMO RIRs obtained with an Eigenmike 32.

As this work only focus on the localisation the phantom sources auralised with an upscaled SRD-RIR, it may be interesting to investigate other spatial acoustics criteria and more generally its fidelity of restitution toward a real environment, compared to more common RIR auralisation techniques.

Nomenclature

ϕ_S azimutal beam direction at the source

ϕ_{head} azimutal direction of the head

$\boldsymbol{\theta}_l$ normed position vector of the l-th loudspeaker

$\boldsymbol{\Theta}_R$ Direction Of Arrival (DOA)

$\boldsymbol{\Theta}_S$ Direction Of Departure (DOD)

$\boldsymbol{\theta}_S$ beam direction at the source

$\boldsymbol{\theta}_{head-tracking}$ tracked direction of the listener's head

$\boldsymbol{\vartheta}_R$ variable unit vector at the receiver

$\boldsymbol{\vartheta}_S$ variable unit vector at the source

g_R far field directivity of the receiver

g_S far field directivity of the source

δ Kronecker-delta function

F_T zero phase moving average filter of time span T

F_b zero phase band-pass filter of band index b

$F_{p,N-p}$ inverse F distribution for α-confidence and parameters p and $N-p$

P_n^m associated Legendre polynomial of order n and degree m

Y_n^m real valued spherical harmonic of order n and degree m

h impulse response

$h_\circ$ omni-to-omni impulse response

$h_{\circ,x}$, $h_{\circ,y}$, $h_{\circ,z}$ omni-to-x/y/z-dipole impulse response

$h_{l,\circ}$ l-th loudspeaker-to-omni impulse response

$\mathbf{r}_E$ "energy" vector

f temporal frequency variable

N number of points

N_R maximum order at the receiver

N_S maximum order at the source

p number of dimension

t time variable

References

[fN09] I. O. for Normalization, *ISO 3382-1 Acoustics - Measurement of Room Acoustic Parameters. Part 1 : Performannce Rooms*. ISO, 2009.

[HAD06] R.-M. Hom, V. Algazi, and R. Duda, "High-frequency interpolation for motion-tracked binaural sound," *Audio Engineering Society - 121st Convention Papers 2006*, vol. 3, pp. 1166–1178, 01 2006.

[JHN10] D. P. Jarret, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *proc. European Signal Porcessing Conference (EUSIPCO)*, Aalborg, Denmark, August 2010, pp. 442–446.

[JW07] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, sixth edition ed. Pearson Prentice Hall, 2007, ch. 5.

[Lös13] S. Lösler, "Schallfeldspezische Entzerrung bei Radialfiltern begrenzter Dynamik für das Eigenmike," Master's thesis, Institute of Electronic Music and Acoustics, May 2013.

[LR10] A. Lindau and S. Roos, "Perceptual evaluation of discretization and interpolation for motion-tracked binaural (mtb) recordings," *26. Tonmeistertagung - VDT International Convention*, pp. 680–701, November 2010.

[PKDV13] M. Pollow, J. Klein, P. Dietrich, and M. Vorländer, "Including directivity patterns in room acoustical measurements," *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, p. 015008, 2013. [Online]. Available: https://asa.scitation.org/doi/abs/10.1121/1.4800303

[PM05] V. Pulkki and J. Merimaa, "Spatial impulse response rendering: A tool for reproducing room acoustics for multi-channel listening," *Journal of the Audio Engineering Society*, vol. 53, pp. 1115–1127, 12 2005.

[Raf15] B. Rafaely, *Fundamentals of Spherical Array Processing*, ser. Springer Topics in Signal Processing. Springer Berlin Heidelberg, 2015.

[RBF+17] M. Romanov, P. Berghold, M. Frank, D. Rudrich, M. Zaunschirm, and F. Zotter, "Implementation and evaluation of a low-cost headtracker for binaural synthesis," in *Audio Engineering Society Convention 142*, May 2017. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18567

[SZZ18] F. Schultz, M. Zaunschirm, and F. Zotter, "Directivity and electro-acoustic measurements of the io," in *Audio Engineering Society Convention 144*, May 2018.

[TP15] S. Tervo and A. Politis, "Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1539–1551, Oct 2015. [Online]. Available: http://dx.doi.org/10.1109/TASLP.2015.2439573

[TPKL13] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc*, vol. 61, no. 1/2,

pp. 17–28, 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16664

[Uns27] A. Unsöld, "Beiträge zur Quantenmechanik der Atome," *Annalen der Physik*, 1927.

[WSF+17] F. Wendt, G. K. Sharma, M. Frank, F. Zotter, and R. Höldrich, "Perception of spatial sound phenomena created by the icosahedral loudspeaker," *Computer Music Journal (special issue HDLA)*, 2017.

[ZBS+17] M. Zaunschirm, C. Baumgartner, C. Schörkhuber, M. Frank, and F. Zotter, "An efficient source-and-receiver-directional rir measurement method," 2017.

[ZF12] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *Journal of Audio Engineering Society*, vol. 60, no. 10, p. 807, October 2012.

[ZF15] ——, "Investigation of auditory objects caused by directional sources in rooms," *ACTA PHYSICA POLONICA A*, 2015.

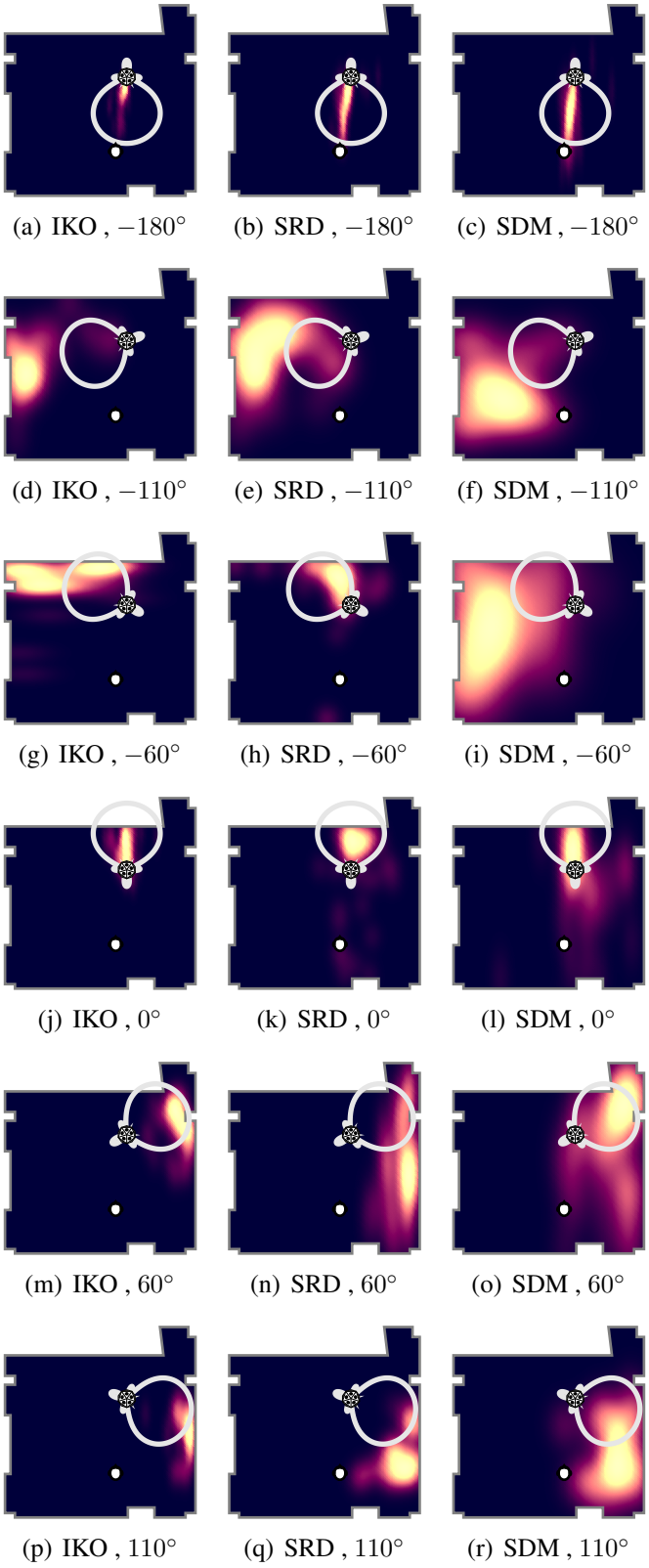
[ZFZ18] M. Zaunschirm, M. Frank, and F. Zotter, "BRIR synthesis using first-order microphone arrays," in *Audio Engineering Society Convention 144*, 2018.

[ZSH18] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 06 2018.

[ZZF07] C. Zhang, Z. Zhang, and D. Florêncio, "Maximum likelihood sound source localization for multiple directional microphones," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007.*, vol. 1. IEEE, 2007, pp. I–125.

[ZZFK17] F. Zotter, M. Zaunschirm, M. Frank, and M. Kronlachner, "A beamformer to play with wall reflections: The icosahedral loudspeaker," *Computer music journal*, vol. 41, no. 3, pp. 50–68, 11 2017.

A PDF estimate for experiment I



(a) IKO , $-180°$ (b) SRD , $-180°$ (c) SDM , $-180°$

(d) IKO , $-110°$ (e) SRD , $-110°$ (f) SDM , $-110°$

(g) IKO , $-60°$ (h) SRD , $-60°$ (i) SDM , $-60°$

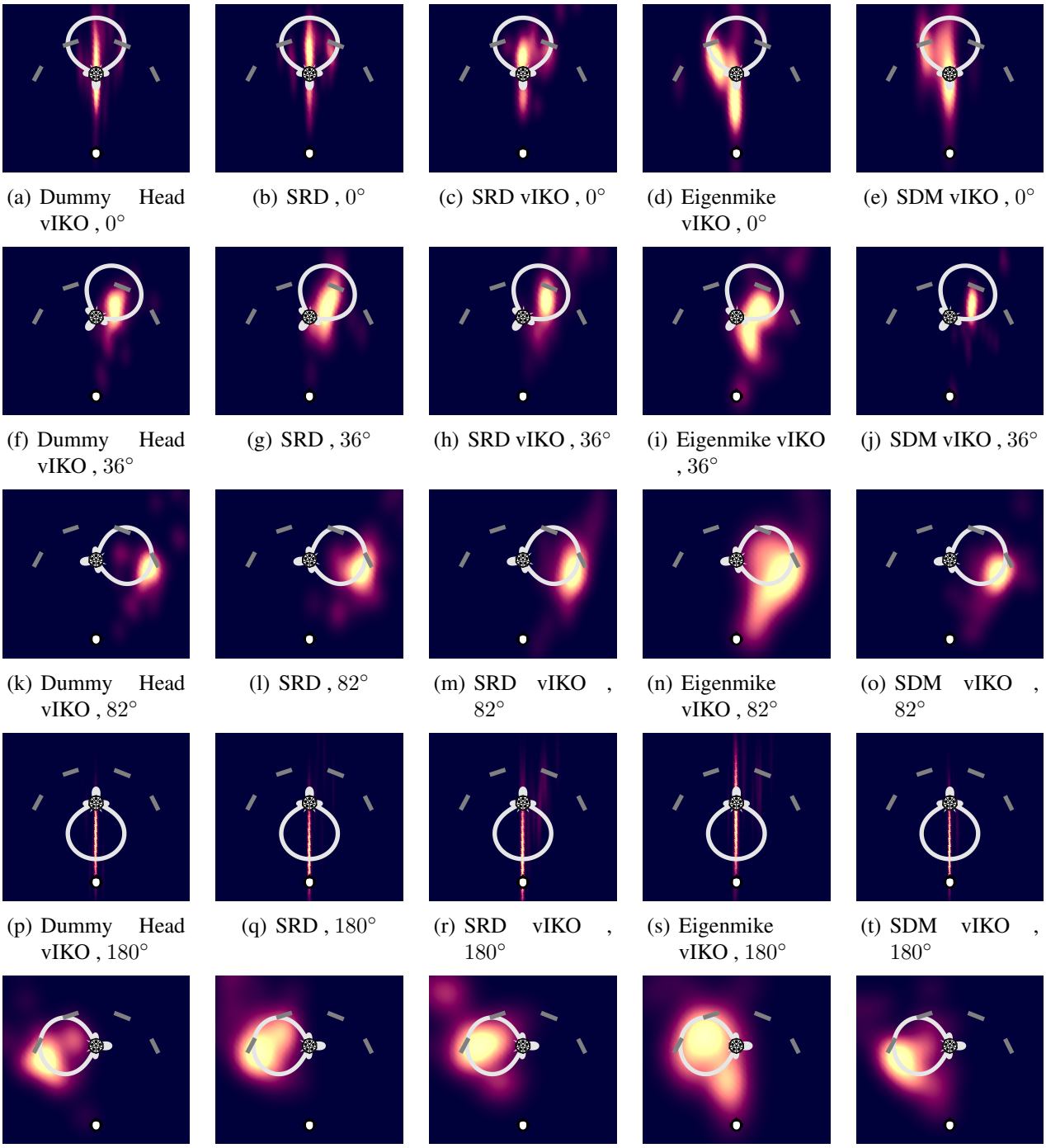
(j) IKO , $0°$ (k) SRD , $0°$ (l) SDM , $0°$

(m) IKO , $60°$ (n) SRD , $60°$ (o) SDM , $60°$

(p) IKO , $110°$ (q) SRD , $110°$ (r) SDM , $110°$

Figure 13 – Probability density function estimates using Kernel density estimation. The given angles correspond to the beam direction ϕ_S

B PDF estimate for experiment II



(a) Dummy Head vIKO , 0° (b) SRD , 0° (c) SRD vIKO , 0° (d) Eigenmike vIKO , 0° (e) SDM vIKO , 0°

(f) Dummy Head vIKO , 36° (g) SRD , 36° (h) SRD vIKO , 36° (i) Eigenmike vIKO , 36° (j) SDM vIKO , 36°

(k) Dummy Head vIKO , 82° (l) SRD , 82° (m) SRD vIKO , 82° (n) Eigenmike vIKO , 82° (o) SDM vIKO , 82°

(p) Dummy Head vIKO , 180° (q) SRD , 180° (r) SRD vIKO , 180° (s) Eigenmike vIKO , 180° (t) SDM vIKO , 180°

(u) Dummy Head vIKO , $-90°$ (v) SRD , $-90°$ (w) SRD vIKO , $-90°$ (x) Eigenmike vIKO , $-90°$ (y) SDM vIKO , $-90°$

Figure 14 – Probability density function estimates using Kernel density estimation. The given angles correspond to the beam direction ϕ_S