Efficient prediction of the listening area for plausible reproduction

Master Thesis

BSc Eric Kurz

Supervisor: DI Ph.D. Matthias Frank
Assessor: Univ.Prof. Dipl.-Ing. Dr.techn. Alois Sontacchi

Graz, November 6, 2018

institut für elektronische musik und akustik

EIDESSTATTLICHE ERKLÄRUNG [1]

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ....
(Unterschrift)

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, the ....
(signature)

1. Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008. Genehmigung des Senates am 01.12.2008

Danksagung

An dieser Stelle will ich mich bei all denjenigen bedanken, die mich während dem Erstellen der nun vorliegenden Masterarbeit unterstützt und begleitet haben. Zunächst danke ich meinem Betreuer DI Ph.D. Matthias Frank, der die Anregung zur Bearbeitung der Thematik sowie hilfreiche Denkanstöße gab und konstruktive Kritik bei der Erstellung der Arbeit übte. Ich danke auch für die Bereitstellung eines Arbeitsplatzes am Institut für Elektronische Musik und Akustik (IEM) während der gesamten Dauer der Masterarbeit. Ebenfalls möchte ich mich beim Kollegium Büros der Petersgasse 116 für die gemeinsam verbrachten Stunden, die anregenden Gespräche und die Hilfsbereitschaft bedanken. Ein besonderer Dank gilt den Teilnehmern und Teilnehmerinnen des durchgeführten Hörversuchs, ohne die diese Arbeit nicht hätte entstehen können. Weiterhin danke ich Mag. Angelika Zuber für die Hilfe bei der Lösung des ein oder anderen mathematischen Problems. Abschließend möchte ich mich bei meinen Eltern bedanken, die mir mein Studium durch ihre Unterstützung ermöglicht haben und mir bei Problemen und Sorgen zur Seite standen.

Zusammenfassung

Bei Planung und Installation umgebender Lautsprecheranordnungen zur Reproduktion dreidimensionaler Schallfelder ist u.a. die Größe des bespielbaren Publikumsbereichs mit plausibler Lokalisation von großer Bedeutung. Dieser Bereich, kurz Hörbereich genannt, wird heutzutage auf Grund von Erfahrungswerten während der Planung geschätzt und kann jedoch erst bei der Inbetriebnahme einer Lautsprecheranordnung vollständig bestimmt werden. Um eventuelle Planungsfehler zu vermeiden ist die Prädiktion des Hörbereichs von großer Wichtigkeit. Bei der Prädiktion des Hörbereichs hat sich die Verwendung eines effizienten erweiterten Energievektors als sinnvoll erwiesen [43]. Die Erweiterung des Energievektors versucht Laufzeit- und Pegelunterschiede in Abhängigkeit der Hörposition zu berücksichtigen. Ausgehend von Erkenntnissen aus [43] wird in dieser Masterarbeit ein existierender Algorithmus zur Prädiktion des Hörbereichs weiter entwickelt. Es wird ein frequenzabhängiges Panning anhand der Ergebnisse von [37] integriert. Weiterhin wird ein Spiegelquellenmodell implementiert um Wandreflexionen beliebiger Ordnung simulieren zu können. Außerdem wird die Abstrahlcharakteristik der verwendeten Lautsprecher integriert. Hierfür werden die Abstrahlcharakteristiken verschiedener Lautsprechermodelle aufgenommen. Es wird ein Hörversuch geplant und durchgeführt, anhand dessen bestimmte Parameter des bestehenden Algorithmus optimiert werden.

Abstract

For planning and installation of surrounding loudspeaker arrays for reproduction of three-dimensional sound fields, among other objectives, the size of the auditorium with plausible localization is quite relevant. Nowadays this area, also called listening area, is estimated during the planning phase. This estimation is based on experience knowledge. However, the size of the listening area cannot exactly be determined before the loudspeaker array is installed. The prediction of the listening area is very important to avoid possible mistakes in the planning process. For prediction of the listening area usage of an efficient extended energy vector is quite reasonable [43]. The extension of the energy vector considers time and level differences in dependence on the listening position. Proceeding from [43], this master thesis refines a prediction algorithm. Based on results from [37], frequency-dependent panning is integrated. Moreover a model of image-sources is implemented to simulate wall reflections of arbitrary order. Furthermore the radiation pattern of the loudspeakers in the surrounding array are integrated. Therefore radiation patterns of different loudspeakers are measured. A listening experiment is performed to optimize parameters of the algorithm.

Contents

List of Figures

List of Tables

1 Introduction

For reproduction of an immersive soundscape for a large audience, we can draw on different playback techniques nowadays. There are channel-based methods (e.g. Auro-3D or Vector-Base Amplitude Panning (VBAP) [56]), object-based methods (e.g. wave field synthesis [1,2] or Dolby Atmos) and scene-based methods (e.g. Higher Order Ambisonics (HOA) [30]). Channel-based methods need a standardized loudspeaker (LS) array for playback, which is defined in the ITU-R BS.2051 [39] for arrays up to 22 LSs. Channel-based methods are very often used for playback systems in cinema halls or at home. A disadvantage of this methods is a relatively small sweet spot around the center of the LS array. Moreover a standardized layout of the LS array is needed. Wave field synthesis can reproduce arbitrary wave fronts physically correct below a certain critical frequency. The method is based on Huygens' principle, which states that arbitrary wave fronts can be represented as the superposition of many so-called elementary waves [3]. With this method the position of the virtual sound source is not limited on the enveloping surface of the LS array. It can be placed in front and behind the array. The listening area reaches over the whole area in front of the array. To have an exact reproduction of a sound field in the full auditory range and for a big auditory area, a tremendous amount of LSs is required. The scene-based method HOA needs substantially less LSs for playback. Moreover LSs of a HOA array can be placed arbitrarily as long as the enveloping surface of the LS dome is evenly sampled. For reproduction, a sound field is decomposed into spherical harmonics (SH) on the enveloping surface. For playback this SH sampled on the LS positions. Similar to channel-based methods, the area, where physically exact reproduction is possible, is around the center position. At playback with 1st-order Ambisonics at 700 Hz the physical sweet spot is equal to the size of the head of a single listener at the center position [72].

In contrast, Frank and Zotter could show that the sweet spot for plausible localisation is substantially bigger [26]. Consequently, theory of sound field reproduction with HOA and human auditory perception differ from each other. In consequence of this fact the auditory area of HOA LS arrays has to be measured via time-consuming listening tests, especially when the influence of different parameters (e.g. decoder weighting, delay compensation, order of playback, etc.) should be considered. For prediction of localization and source width of a virtual sound source at HOA playback, good results can be achieved by the energy vector ($\mathbf{r_E}$-vector) [22–24]. Furthermore, it could be showed that length changes of the $\mathbf{r_E}$-vector during panning (e.g. with VBAP) of a virtual sound source describe changes in timbre. Also for prediction of localization at off-center listening positions the $\mathbf{r_E}$-vector was proposed in [11]. An extended $\mathbf{r_E}$-vector model was developed by Stitt [69]. Stitt tried to integrate the precedence effect as well as the „Cone-of-Confusion" into the vector model. Unfortunately, this extensions result in a quite high computational expense. Through a performance comparison of different $\mathbf{r_E}$-vector model extensions in the previous work [43] it could be shown that the high computational effort of Stitt's model does not yield significantly better results in prediction of sound source localization. A quite simpler but equally performing vector model extension was introduced in [43] (cf. eq. 5). This extension tries to consider

sound dissipation through distance-depended damping weights $w_{d,i}$ and the law of the first wavefront, also called precedence effect, through time-dependent weights $w_{\tau,i}$.

Following the previous work, motivation for this master thesis is the further development of the existing extended energy vector model introduced in [43]. Initially, frequency-dependent panning for virtual sound sources according to [37] is considered in the vector model. Furthermore, a image-source model for simple room geometrics is integrated into the extended $\mathbf{r_E}$-vector to model the influences of sound reflections on the localization. Furthermore direction- and frequency-dependent radiation patterns of the LSs that are used in a spherical array, are included into the model. Therefore, radiation patterns of existing LSs was measured with a circular microphone array. To represent this radiation patterns in a more compact way, they are transformed into the SH domain. The time-dependent weights $w_{\tau,i}$ of the extended model are dependent on the so called echo threshold slope τ, which relates a certain delay of an incoming sound wave with an appropriate level damping. To align w_τ into the context of spatial hearing, a short literature study is performed. It focusses especially on the case of stereophonic playback as the simplest form of spatial hearing with multiple sound sources. Here the connection between interchannel level and time differences (ICLDs and ICTDs) is investigated. For finding suitable w_τ, a listening experiment is planned and performed. The resulting data is used to perform an optimization for τ. Moreover the evaluation of the experiment data focusses on a delay threshold for source splitting effects.

2 Vector models for localization

2.1 Basic vector model

To explain the basic principle of a vector model describing the playback direction of a virtual sound source, we initially focus on stereophonic amplitude panning for horizontal LS pairs (cf. fig. 1).

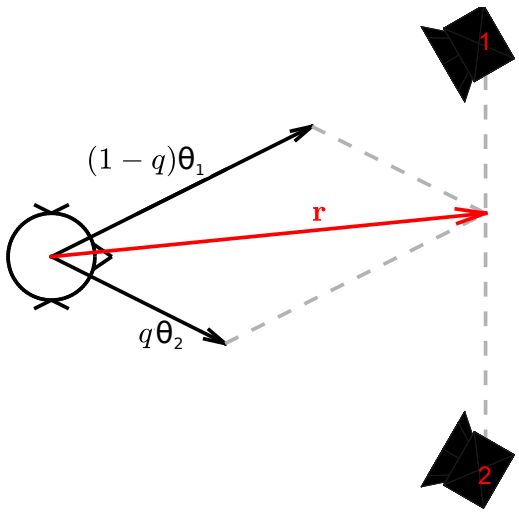


Figure 1 – Principle of a vector model for localization for stereophonic amplitude panning setup. LSs (1,2) and listener in the center position of the setup. Weighted LS direction vectors $(1-q)\boldsymbol{\theta}_1$ and $q\boldsymbol{\theta}_2$ as well as resulting panning direction vector $\mathbf{r}$ (red).

The principle shown in fig. 1 can also be expressed through

$$\mathbf{r} = (1-q)\boldsymbol{\theta}_1 + q\boldsymbol{\theta}_2 \quad \text{with} \quad 0 \leq q \leq 1. \tag{1}$$

q is the blending parameter that linearly blends between LS 1 and 2. If we choose $q = \frac{g_2^2}{g_1^2+g_2^2}$ (normalized gain g_2) we get

$$\mathbf{r_E} = \left(1 - \frac{g_2^2}{g_1^2+g_2^2}\right)\boldsymbol{\theta}_1 + \frac{g_2^2}{g_1^2+g_2^2}\boldsymbol{\theta}_2 = \frac{g_1^2\boldsymbol{\theta}_1 + g_2^2\boldsymbol{\theta}_2}{g_1^2+g_2^2} = \frac{\sum_{i=1}^{2} g_i^2\boldsymbol{\theta}_i}{\sum_{i=1}^{2} g_i^2} \tag{2}$$

whereby g_1 and g_2 are the gains for the LSs. This principle can be applied to an arbitrary amount of LSs that are facing on a center position. Consequently the $\mathbf{r_E}$-vector can be seen as the superposition of all sound energy vectors arriving at a receiving point and it can be computed according to [33]

$$\mathbf{r_E} = \frac{\sum_{i=1}^{I} g_i^2 \boldsymbol{\theta}_i}{\sum_{i=1}^{I} g_i^2}, \tag{3}$$

whereby variables are defined as follows:

g_i ... weigth of the i-th LS
$\boldsymbol{\theta}_i$ ... unit vector in main radiation direction of the i-th LS.

According to Gerzon [33] the $\mathbf{r_E}$-vector is a second degree first-order model, which considers the influence of head movements.

As investigated in [22, 24, 28] the energy vector model is a good predictor for localization at the center listening position for frontal and horizontal surrounding playback situations for broad band sound sources. For amplitude panning it could be shown that the $\mathbf{r_E}$-vector outperforms binaural models, e.g. regarding to Jeffress [40], Dietz [15] or Lindemann [47, 48], when it comes to localization prediction.

The normalization in eq. 3 is done to limit the magnitude of $||\mathbf{r_E}||$ between 0 (sound from everywhere resp. from opposing directions) and 1 (sound from a single direction). Every phantom sound source that is produced by more than one LS obviously, leads to a magnitude of $||\mathbf{r_E}|| < 1$ per definition. For playback with Ambisonics, it is desirable that the magnitude of th $\mathbf{r_E}$-vector is as large as possible and constant for all panning directions. This criterion is used for the design of the so called max-$\mathbf{r_E}$ decoder (cf. [32]).

Moreover the perceived source width can be described with the help of the $\mathbf{r_E}$-vector [23]. For frontal phantom sound sources W is indirect proportional to the cosine of the magnitude of the vector (cf. 4). When a plane with distance $||\mathbf{r_E}||$ from the origin cuts a unit sphere the expression $2 \arccos ||\mathbf{r_E}||$ describes the aperture angle of this cap cut-off. The cut-off corresponds to the size of the LS pair used to playback the stimuli in the listening experiment of [23]. It was figured out that listeners tend to hear a source width of $\frac{5}{8}$ of this aperture angle

$$W = \frac{5}{8} \cdot \frac{180^\circ}{\pi} \cdot 2 \arccos ||\mathbf{r_E}||. \tag{4}$$

2.2 Existing extended energy vector model

The energy vector model, which is used as a base for all upcoming extensions is described through

$$\mathbf{r_E} = \frac{\sum_{i=1}^{I} (w_{\tau,i} w_{d,i} g_i)^2 \boldsymbol{\theta}_i}{\sum_{i=1}^{I} (w_{\tau,i} w_{d,i} g_i)^2}. \tag{5}$$

In eq. 5 the $\mathbf{r_E}$-vector model from eq. 3 is extended with sound dissipation weights $w_{d,i}$ and time-dependent weights $w_{\tau,i}$. For modelling of the sound dissipation while a sound wave is propagating through the air, a damping of -6dB per distance doubling is implemented. Weight $w_{d,i}$ for the i-th LS is determined through the quotient

$$w_{d,i} = \frac{1}{||\mathbf{r}_{\boldsymbol{\theta}_{1\ldots i},\mathbf{v}}||}. \tag{6}$$

$\mathbf{r}_{\boldsymbol{\theta}_{1\ldots i},\mathbf{v}} = \mathbf{v} - \boldsymbol{\theta}_{1\ldots i}$ stands for vectors, which are tensed between the listening position and LSs 1 to i.

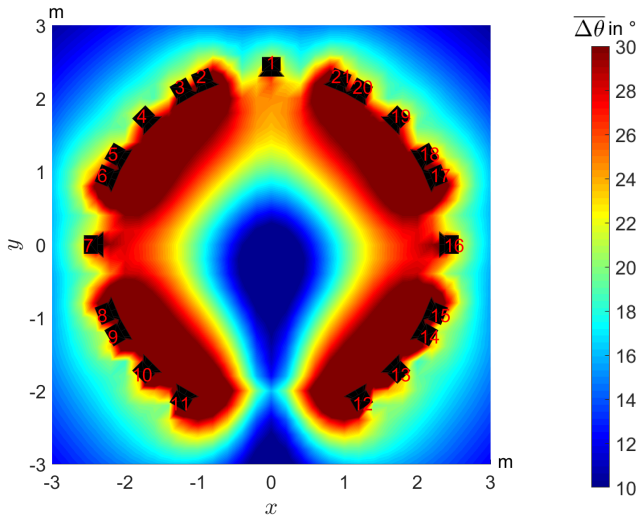
A quite simple approach to consider the law of the first wave front is to give sound waves, arriving delayed at the listening position, less weight in computation of the $\mathbf{r_E}$-vector. Sound waves are sorted ascending by there arrival times T_1 to T_i, where a larger arrival time T_i stands for a larger delay. The delay of the i-th signal is

$$\Delta t_i = T_i - T_1 \quad \text{with} \quad T_1, T_i \text{ in ms.} \tag{7}$$

Weights $w_{\tau,i}$ can be computed with the echo threshold slope parameter τ = -0.25dB/ms in the following way:

$$w_{\tau,i} = \tau \cdot \Delta t_i. \tag{8}$$

Figure 2 shows the simulation of the mean localization error $\overline{\Delta\theta}$ for a circular LS array as it was used in the listening experiment (cf. section 4). To plot this figure a localization error $\Delta\theta$ is computed for every panning direction θ (1°-steps) at every listening position (grid size is 0.2 x 0.2 m) with the extended energy vector model from eq. 5. Afterwards, the mean value for all $\Delta\theta$ is computed at every listening position and plotted into the figure using the MATLAB-function `surf`. This simulation is done for a playback with 5th-order of Ambisonics with max-$\mathbf{r_E}$ weighting and free-field conditions. A gain or delay compensation is not done because of the circular positioning of the LSs. Moreover the LSs are modelled with omnidirectional radiation patterns. The sweetspot for localization is clearly visible in the center of the array. This sweetspot has a lightbulb-like shape because of the missing LS in the back of the array. The more the listening position moves to the LSs the worse gets localization.



(a) (b)

Figure 2 – Mean localization error $\overline{\Delta\theta}$ in dependence on the listening position for 5th-order Ambisonics and max-$\mathbf{r_E}$ weighting with free-field conditions. No gain and delay compensation is done. LSs are modelled with omnidirectional radiation patterns.

2.3 Frequency-dependent panning

Previous investigations stated that it would make sense to introduce an adjustable slope parameter γ to the vector model [28, 76]. Considering this, it is possible to generalize eq. 3 to

$$\mathbf{r}_\gamma = \frac{\sum_{i=1}^{I} |g_i|^\gamma \boldsymbol{\theta}_i}{\sum_{i=1}^{I} |g_i|^\gamma}. \tag{9}$$

Consequently $\mathbf{r_E}$ is the special case of eq. 9 for $\gamma = 2$. Of course eq. 9 can be written down for the stereophonic case (cf. fig. 1).

$$\mathbf{r}_\gamma = \frac{|g_1|^\gamma \boldsymbol{\theta}_1 + |g_2|^\gamma \boldsymbol{\theta}_2}{|g_1|^\gamma + |g_2|^\gamma}. \tag{10}$$

This equation is equivalent to the generalized stereophonic tangent law that was proposed in [84].

$$\frac{\tan\varphi}{\tan\alpha} = \tanh\left(\gamma \cdot \frac{ln10}{40}(\Delta L - W)\right) \tag{11}$$

Whereby variables are defined as follows:

φ	...	perceived lateralisation of the phantom sound source in °
α	...	half aperture angle of the LS pair in°
γ	...	slope of the amplitude-panning curve at the center position
ΔL	...	level difference between the LSs in dB

W ... horizontal shift of the amplitude-panning curve in dB

Figure 3 shows panning curves in dependence on γ for the listening setup that is shown in fig. 1.

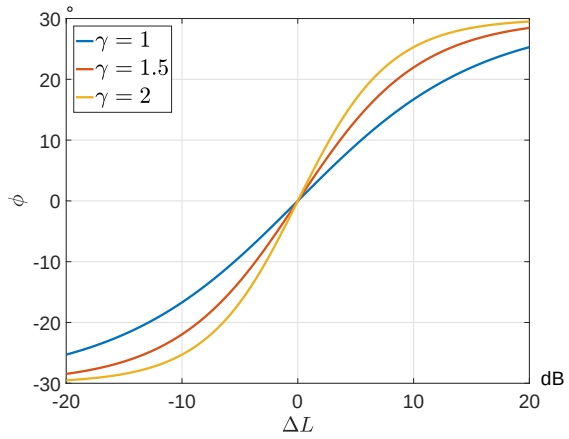


Figure 3 – Panning curves according to eq. 11 for different slope parameters γ.

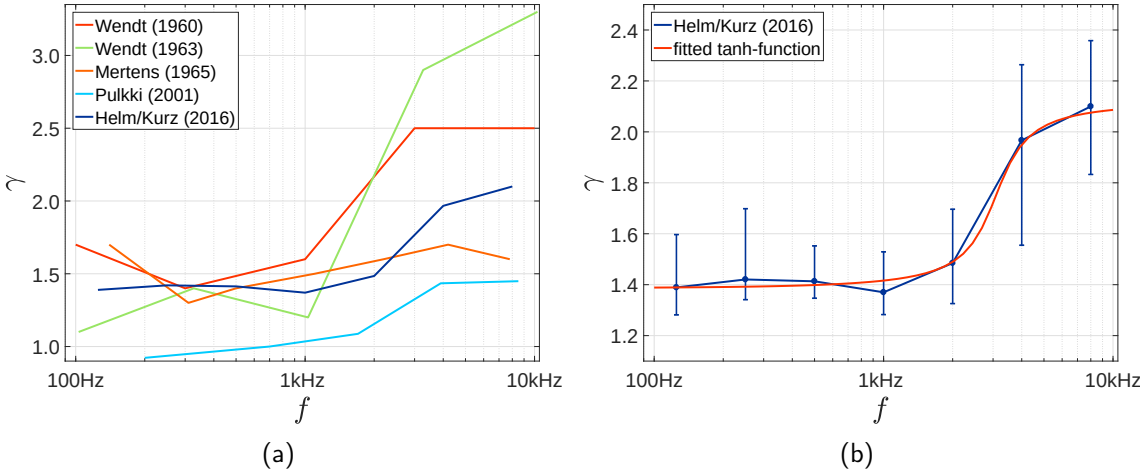


Figure 4 – (a): γ values in dependency on the frequency from various investigations. (b): Data from [37]. Determined γ values with corresponding 95% confidence intervals and interpolated curve (blue). Fitted $\tanh$-function for $\gamma(f)$ (red).

In previous investigations about localization for stereophonic playback, it could be shown that the slope parameter γ is frequency-dependent [54, 57, 74, 75]. Also Helm and the author himself did some research about this topic [37]. Figure 4 (a) shows the results of this investigation. For $f \leq$ 1kHz the frequency-dependent $\gamma(f)$ from [37] mostly in the range of the curves from Mertens (1965) and Wendt (1960 and 1963). Interestingly, the curve of Pulkki (2001) shows a lower slope with values even below 1 for $f \leq$ 700Hz. For higher frequencies ($f \geq$ 2kHz) the curves of Wendt tend to higher values for $\gamma(f)$ whereas Mertens and Pulkki tend towards $\gamma(f) = 1.5$. Results from Helm/Kurz lie

between these curves. In general, the curve from Helm/Kurz shows a simpler and more monotone slope for $\gamma(f)$ which can be divided into 3 sections. A constant section in the lower range up to 1 kHz, a transition stage in the mid-frequency range from 1 to 3 kHz and a section with a sleight slope of $\gamma(f)$ for frequencies higher than 3 kHz. Figure 4 (b) shows the same resulting progression of the slope parameter γ with median values as supporting points and corresponding 95% confidence intervals (blue) for frequencies from 100 Hz up to 10 kHz. Also a fitted curve for $\gamma(f)$, which was computed on basis of the median values, is shown in fig. 4 (b) (red). The corresponding fit-function is

$$\gamma(f) = 0.252 \cdot \tanh\left(\frac{f}{804\ \text{Hz}} - 3.677\right) + 1.715. \tag{12}$$

In order to limit the computational effort for this extension of the $\mathbf{r_E}$-vector model it was decided to divide this curve in 3 bands and compute mean values for γ (cf. tab. 1).

100Hz $\leq f \leq$ 1kHz	1kHz $\leq f \leq$ 5kHz	$f \geq$ 5kHz
$\overline{\gamma}_1 = 1.464$	$\overline{\gamma}_2 = 1.637$	$\overline{\gamma}_3 = 1.967$

Table 1 – Frequencies bands and depending values for $\overline{\gamma}_j$.

Consequently the resulting $\mathbf{r_E}$-vector model is

$$\mathbf{r_E} = \frac{\sum_{i=1}^{I} |w_{\tau,i} w_{d,i} g_i|^{\overline{\gamma}_j} \boldsymbol{\theta}_i}{\sum_{i=1}^{I} |w_{\tau,i} w_{d,i} g_i|^{\overline{\gamma}_j}}. \tag{13}$$

Figures 5 (a), (b) and (c) are showing the simulation for $\overline{\Delta\theta}$ according to eq. 13 with $\overline{\gamma}_1$, $\overline{\gamma}_2$, and $\overline{\gamma}_3$ from table 1 for the corresponding frequency bands. The simulation is done under the same conditions as in fig. 2. It is visible that for an increasing slope γ the mean localization error $\overline{\Delta\theta}$ becomes higher in the peripheral section of the listening area. Moreover when moving to off-center positions, localization becomes worse much faster for higher values of γ. This simulation result coincides well with the observations made in [37, 54, 57, 74, 75] that for higher frequencies, human hearing is more sensitive for ICLDs. In other words, localization is better for higher frequencies and therefore imaging errors of the virtual have more influence on the listeners perception.

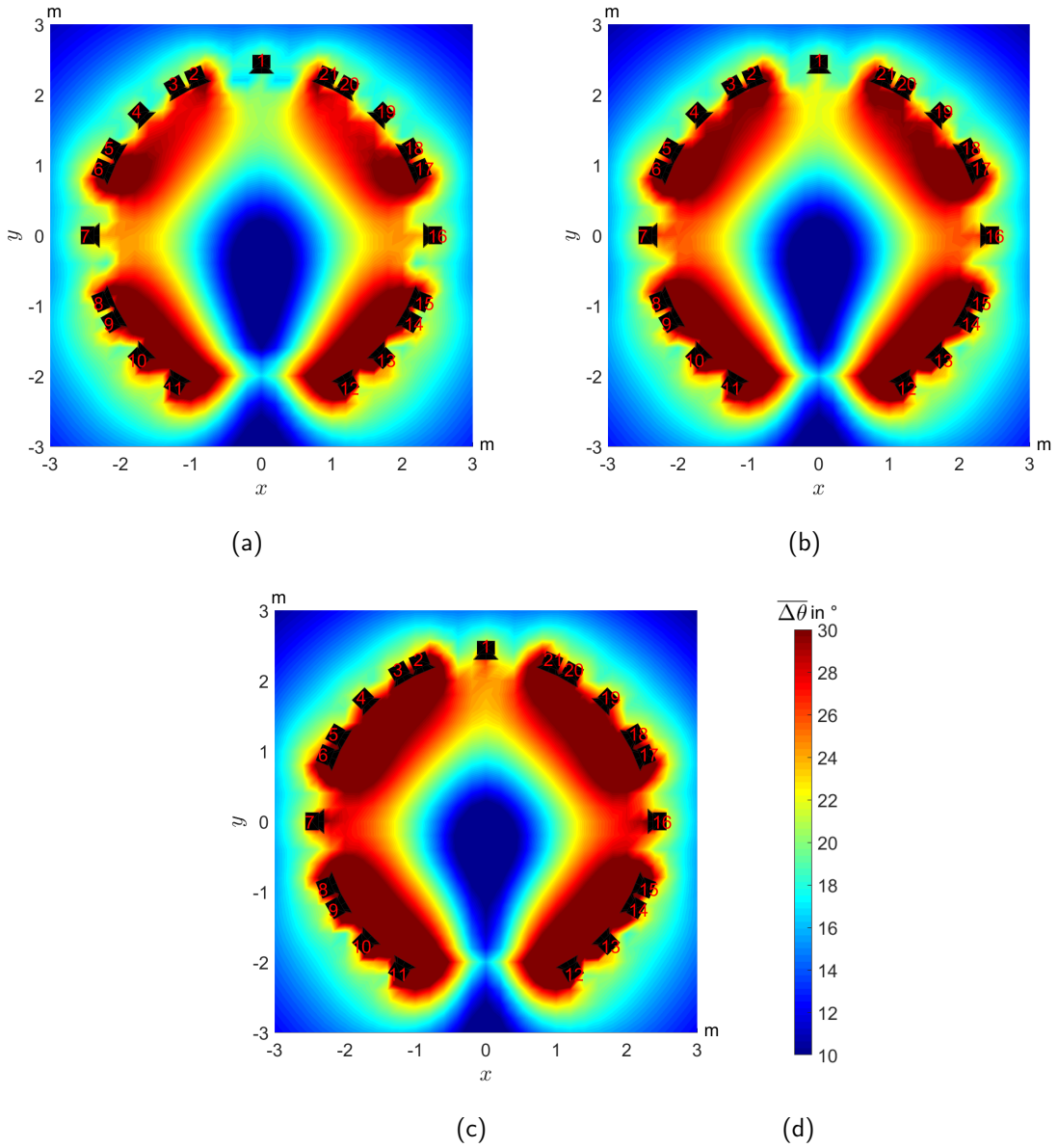


Figure 5 – Mean localization error $\overline{\Delta\theta}$ in dependence on the listening position for 5th-order Ambisonics and max-$\mathbf{r_E}$ weighting with free-field conditions. No gain and delay compensation is done. LSs are modelled with omnidirectional radiation patterns. Simulation is done with a slope of $\overline{\gamma}_1 = 1.464$ (a), $\overline{\gamma}_2 = 1.637$ (b) and $\overline{\gamma}_3 = 1.967$ (c) according to the frequency bands in tab. 1.

2.4 Image-source model

In order to investigate influences from room acoustics, more precisely from reflections, on the localization, a simple n-th order three-dimensional image-source model was integrated in the $\mathbf{r_E}$ algorithm. The image-source model can be developed from geometrical room acoustics. Basic principle of geometrical room acoustics is the assumption that sound propagates ray-like inside a room. These sound rays can be seen equivalent to light rays and principles from geometrical optics can be used to describe sound propagation. Especially at large and plain boundary surfaces of the room, i.e. large overall dimension and small roughness in comparison to the wavelength, sound reflection according to figure 6 (solid line) can be assumed. We can see that the incidence angle κ is equal to the reflection angle. The resulting sound field can easily be modeled as superposition of the field of the sound source and the field of a mirrored version of this sound source under free field conditions (cf. fig. 6). Length and direction of the sound propagation paths at the receiver position are for both cases remain the same.

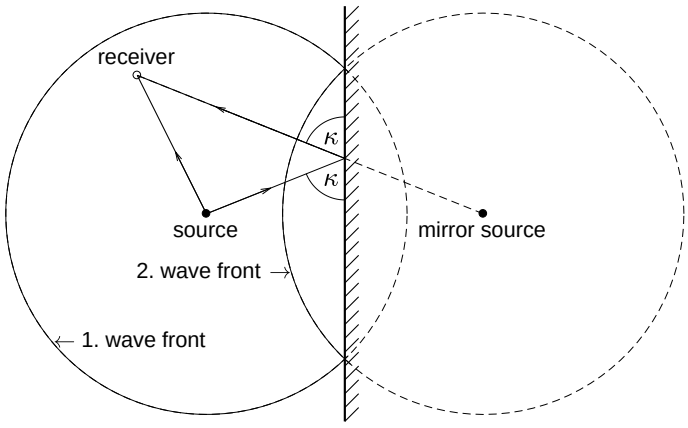


Figure 6 – Principle of sound reflection according to geometrical room acoustics and simulation with a image-source model.

Moreover it is possible to model sound absorption at the boundary surface of the room. The gain factor of a K-th order image-source $g_{i,MS}$ can be easily computed with the gain of the original source g_i and the mean absorption coefficient $\overline{\alpha}$ via

$$g_{i,MS} = g_i \cdot (1 - \overline{\alpha})^{K/2}. \tag{14}$$

In this simple image-source model $\overline{\alpha}$ is assumed as a mean absorption coefficient for all boundary surfaces. It can be derived from the reverberation time T_{60} by simply transposing the equation for T_{60} by Sabine [63] or Eyring. Whereas T_{60} by Sabine (eq. 15) is suitable for rooms with a small absorption coefficient $\overline{\alpha} \leq 0.3$ [42]. For rooms with a $\overline{\alpha} > 0.3$ eq. 17 by Eyring should be used [29].

• T_{60} by Sabine:

$$T_{60} = 0.161 \frac{\mathsf{s}}{\mathsf{m}} \cdot \frac{V}{S \cdot \overline{\alpha} + 4 \cdot m \cdot V} \tag{15}$$

$$\vdots$$

$$\overline{\alpha} = 0.161 \frac{V}{S \cdot T_{60}} - \frac{4 \cdot m \cdot V}{S} \tag{16}$$

The influence of damping m by sound propagation in air is considered by the term $4 \cdot m \cdot V$. Room volume V and area of the boundary surfaces is computed in the following way:

$$V = L_x \cdot L_y \cdot L_z,$$
$$S = 2L_xL_y + 2L_xL_z + 2L_yL_z.$$

• T_{60} by Eyring:

$$T_{60} = -0.161 \frac{\mathsf{s}}{\mathsf{m}} \cdot \frac{V}{S \cdot \ln(1 - \overline{\alpha})} \tag{17}$$

$$\vdots$$

$$\overline{\alpha} = 1 - \exp\left(-0.161 \cdot \frac{V}{S \cdot T_{60}}\right) \tag{18}$$

If there is no reverberation time T_{60} available for the considered room, an optimal $\hat{T}_{60}$ can be computed via the approximation formula

$$\hat{T}_{60} = 0.25 \cdot \sqrt[3]{\frac{V}{100}}. \tag{19}$$

Of course the image-source model can be extended to the three room dimensions and consequently to all boundary surfaces of a simple room geometry (shoebox model). If we mirror a sound source at every boundary surface of a room and sum up all the incoming sound rays, i.e. 1st reflections, at the receiver position, we get a 1st-order image-source model. Also it is possible to mirror the mirrored sources another time at the mirrored boundary surfaces of the room to get the 2nd reflections and the 2nd-order image-source model. The number of mirror sources is determined through

$$\text{2-D:} \quad L = 2K^2 + 2K, \tag{20}$$

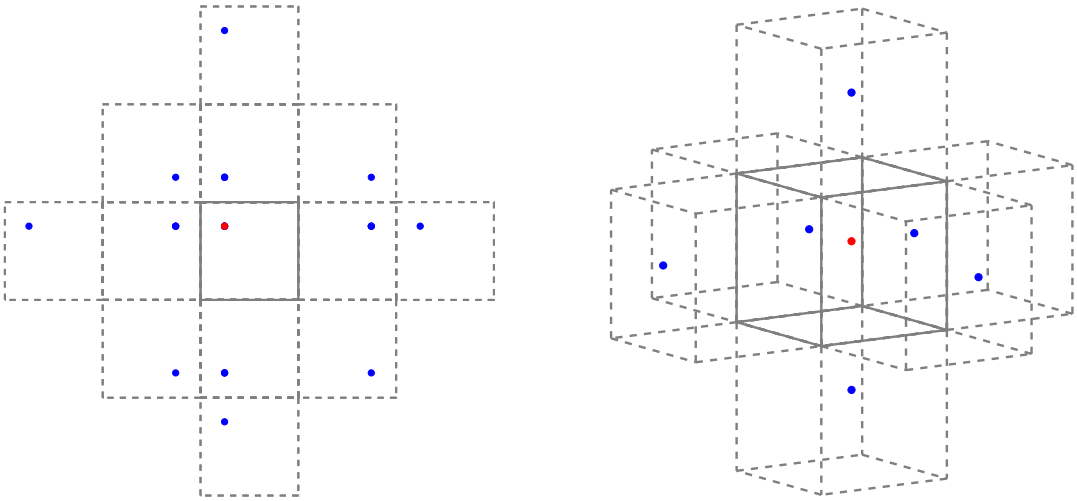
$$\text{3-D:} \quad L = \frac{4}{3}K^3 + 2K^2 + \frac{8}{3}K. \tag{21}$$

Equations 20 and 21 were determined with the polynomial curve fitting function `polyfit` in MATLAB. A more compact way to determine the number of mirror sources is described in eq. 22 and 23 (proof see Appendix eq. 60 and 61).

$$\text{2-D:} \quad L = \sum_{k=1}^{K} 4k, \tag{22}$$

$$\text{3-D:} \quad L = \sum_{k=1}^{K} 4k^2 + 2. \tag{23}$$

Figure 7 shows the position of the mirror sources for a 2-D model (a) and a 3-D model as well as the imaginary rooms for the those sources. To keep the illustration simple a square resp. cube was chosen for the room model and a 2nd resp. 1st-order image-source model was plotted. In both cases the original source (red dot) is at a off-center position to illustrate the principle. This simple model can not be applied to any combination of source-receiver-position and room geometry.



(a) (b)

Figure 7 – (a): 2-D 2nd-order image-source model. (b): 3-D 1st-order image-source model. Original source (red) and mirrored sources (blue).

With equations 14 and 22 resp. 23 the $\mathbf{r_E}$-vector model from eq. 13 can be extended to

$$\mathbf{r_E} = \frac{\sum\limits_{i=1}^{I} \left(|w_{\tau,i} w_{d,i} g_i|^{\overline{\gamma}_j} \boldsymbol{\theta}_i + \sum\limits_{l=1}^{L} |(1-\overline{\alpha})^{K/2} w_{\tau,i,l} w_{d,i,l} g_i|^{\overline{\gamma}_j} \boldsymbol{\theta}_{i,l} \right)}{\sum\limits_{i=1}^{I} \left(|w_{\tau,i} w_{d,i} g_i|^{\overline{\gamma}_j} + \sum\limits_{l=1}^{L} |(1-\overline{\alpha})^{K/2} w_{\tau,i,l} w_{d,i,l} g_i|^{\overline{\gamma}_j} \right)}. \tag{24}$$

Whereas the term $|w_{\tau,i} w_{d,i} g_i|^{\overline{\gamma}_j} \boldsymbol{\theta}_i$ stands for the gain vector of the i-th LS and $\sum\limits_{l=1}^{L} |(1-\overline{\alpha})^{K/2} w_{\tau,i,l} w_{d,i,l} g_i|^{\overline{\gamma}_j} \boldsymbol{\theta}_{i,l}$ represents the gain vectors of its mirrored versions.

Figures 8 (a), (b) and (c) are showing the simulation for the mean localization error $\overline{\Delta\theta}$ considering a 3-D image-source model for a room that is limited by the x- and y-axes of the figures and has a height of $z = 2$m. The order of the image-source model ascends from $K = 1$ (a) to $K = 3$ (c). Again the simulation was performed with 5th-order Ambisonics, max-$\mathbf{r_E}$ weighting, no gain or delay compensation and LSs are modelled as omnidirectional sound sources. First the optimal reverberation time is computed with eq. 19 so that $\hat{T}_{60} = 0.224$s. To compute the mean absorption coefficient eq. 18 is used, so that $\overline{\alpha} = 0.35$. In the next step the positions of the image-sources are computed. Original and image-source positions as well as $\overline{\alpha}$ are deployed to eq. 24. The simulation is performed with a fixed slope of $\overline{\gamma}_j = \gamma = 2$.

When comparing fig. 8 (a) with fig. 2 the influence of the 1st-order image-sources resp. the first reflections is clearly visible. The area with good localization inside the LS array shrinks strongly. Moreover, localization for areas outside the LS ring, which off course are not mainly of interest for our investigations, is worse. Especially when a LS is close to a boundary surface. At simulation with a 2nd-order image-source model (fig. 8 (b)) the area with plausible localization nearly remains constant. For a 3rd-order image-source model (fig. 8 (c)) localization only slightly changes in the center position of the array. Therefore it is sufficient to compute a 2nd-order model. Maybe in special cases a simulation with a 3rd-order model provides more detailed information.

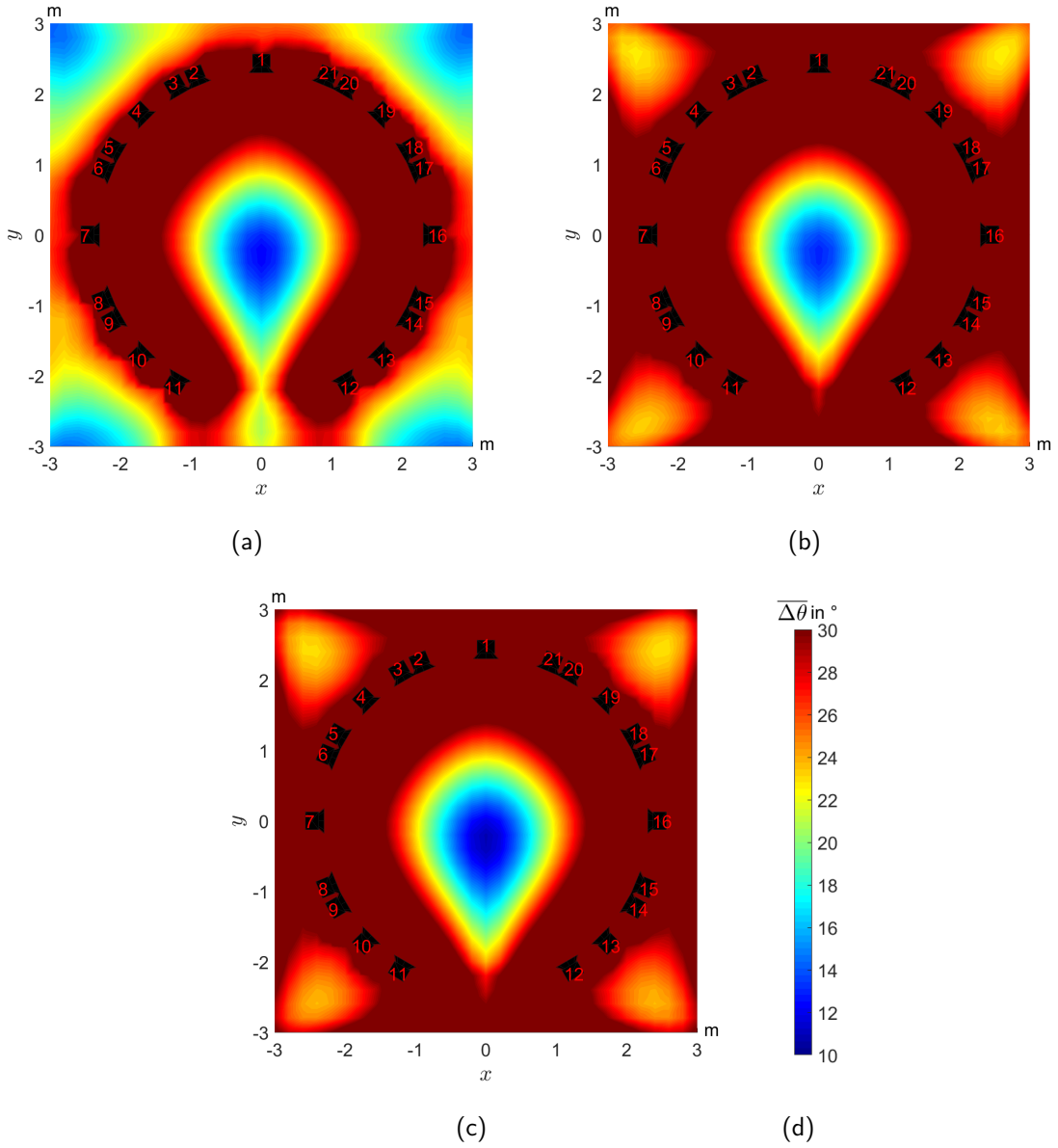


Figure 8 – Mean localization error $\overline{\Delta\theta}$ in dependence on the listening position for 5th-order Ambisonics and max-$\mathbf{r_E}$ weighting with applied 3-D image-source model. No gain and delay compensation is done. LSs are modelled with omnidirectional radiation patterns. $\overline{\alpha} = 0.35$ is computed with eq. 18 and a $\hat{T}_{60} = 0.224$s. The order of the model is (a): $K = 1$, (b): $K = 2$, (c): $K = 3$.

2.5 Loudspeaker radiation patterns

Measurement and analysis of the LS radiation patterns was done referring to the work described in [5].

2.5.1 Setup and measurement

For measurement of LS radiation patterns a circular microphone array out of two wooden circular rings was used (cf. fig. 9). This array was built up in the anechoic measurement chamber of the IEM. One ring was located in the horizontal plane, the other one layed in the vertical plane. The horizontal ring was held by 4 LS stands and the vertical ring was screwed to the horizontal ring at their intersection. Both rings held altogether 61 NTI MA 2230 omnidirectional microphones at a radius $r = 1$m and an angular spacing of 11.25°. 30 microphones were mounted on the horizontal ring and 29 on the vertical ring. Two microphones were mounted exactly at the intersection (front/back) of both rings. The LS, which had to be measured, was mounted on a LS stand at the center position of the setup. Of course the acoustical center of a LS is highly frequency-dependent, but a special positioning to treat these problem was not performed. The LS stand stood on the turntable Outline ET 250-3D [2]. The LS radiation patterns were measured only by the microphones on the vertical ring and those two at the intersection points. The microphone at elevation angle $\theta = 90°$ was neglected. Through rotation of the LS in steps of 10° from 0° to 180° a sampling of the full sphere was possible. The entire set of sampling points results from eq. 25.

$$\begin{aligned} \varphi &= [0° : 10° : 350°] && \Rightarrow \quad N_\varphi = 36 \\ \theta &= [-78.75° : 11.25° : 78.75°] && \Rightarrow \quad N_\theta = 15 \\ &\Downarrow \\ N &= N_\varphi \cdot N_\theta = 540 \end{aligned} \tag{25}$$

The microphones were connected to a microphone preamplifier and AD/DA converter DirectOut Technologies Andiamo.MC [3]. This AD/DA converter was connected via MADI to the PC which ran a Pd patch for measurement controlling (see Appendix fig. 58). Also the turntable was connected via network to the PC and was controlled by the Pd patch. For every rotation step a logarithmic sweep was played back over the LS and the output was recorded via the microphones and the AD/DA converter directly in the Pd patch into two separate 15-channel 44.1kHz 24bit PCM audio files. The first audio file holds the signals from the microphones of the front part of the vertical ring, the second these from the back part.

2. `http://outline.it/outline-products/measurement-systems/et-250-3d/`
3. `https://www.directout.eu/produkte/andiamomc/`

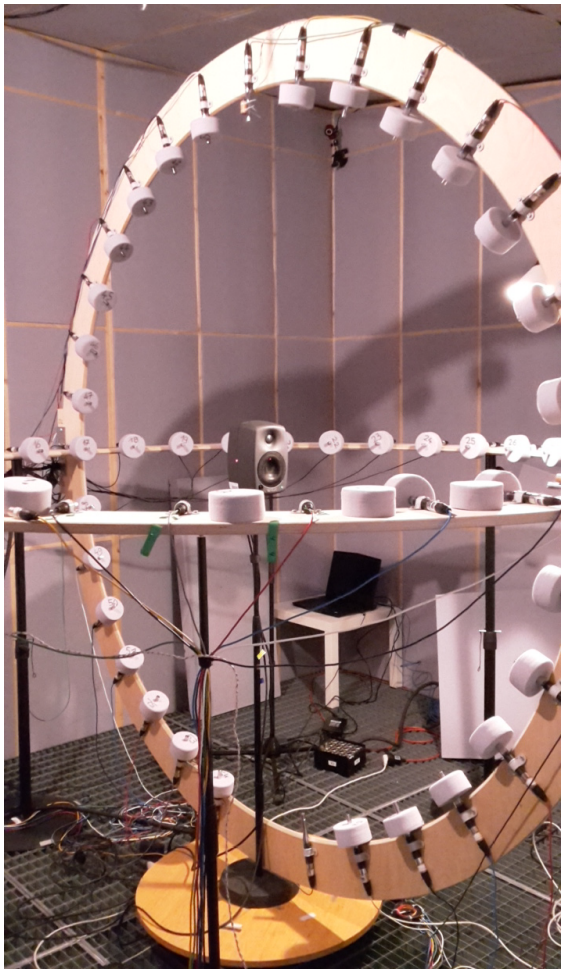


Figure 9 – Circular microphone array with a Genelec 8020 LS inside the anechoic chamber.

2.5.2 Signal processing

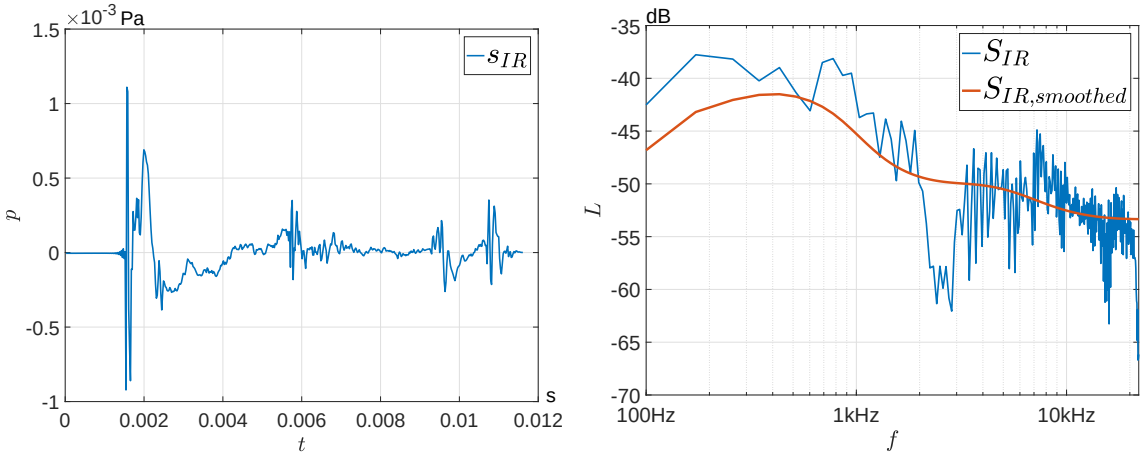
Figure 10 (a) shows the impulse response $p(t)$ of a Genelec 8020 LS on its main radiation direction ($\varphi = \theta = 0°$). This impulse response can be compute from the recorded logarithmic sweep with deconvolution in the frequency domain (cf. eq. 26). To get rid off unwanted room/floor reflections the impulse response is cropped after 512 samples using a half-sided Hann window.

$$s_{IR}(t) = IFFT\left(\frac{FFT(s(t)_{log,rec})}{FFT(s(t)_{log})}\right) \tag{26}$$

In figure 10 (b) we can see the magnitude spectrum $L(f)$ of the impulse response (blue, cf. eq. 27) as well as a three-octave-smoothed version of these (red). The phase spectrum of $s_{IR}(t)$ is neglected in the whole processing.

$$S_{IR}(f) = FFT(s_{IR}(t)) \tag{27}$$

Smoothing was performed with the MATLAB-function `smoothSpectrum`. The function calculates the i-th octave smoothed spectral coefficient as the sum of the Gaussian windowed spectrum around the center frequency $f(i)$. The Gaussian window has a standard deviation of $\sigma = 3 \cdot f(i)$. A smoothed spectrum for every sampling point results in more consistent radiation patterns.



(a) (b)

Figure 10 – (a): Impulse response (sound pressure p in dependence on time t) of a Genelec 8020 LS on its main radiation direction $\varphi = \theta = 0°$. 512 samples recorded with 44.1 kHz. (b): Spectrum of the impulse response (blue). three-octave-smoothed spectrum of the impulse response (red).

Afterwards $S_{IR,smoothed}$ is normalized and linearized and we get the linear gain function $G_{IR}(f)$ (cf. fig. 11). According to the frequency-dependent panning that is outlined

in subsection 2.3, the whole frequency range of $G_{IR}(f)$ is divided into 3 bands: 100 Hz to 1 kHz, 1kHz to 5kHz and 5 kHz to 20 kHz. By taking the mean value of the gains in these bands computational effort for the whole $\mathbf{r_E}$-algorithm is strongly reduced. Figure 11 also shows the mean gains g_1, g_2 and g_3 for the bands.

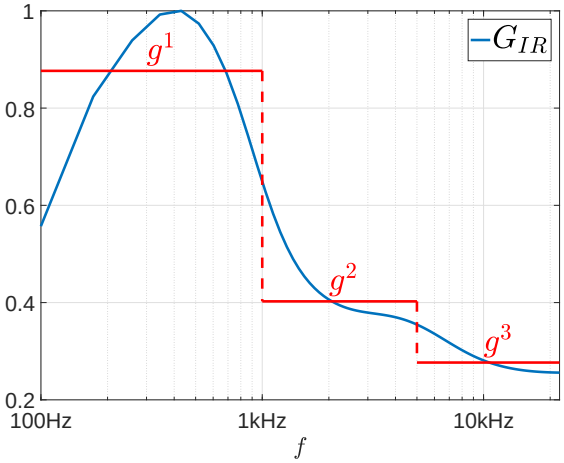


Figure 11 – Linear gain function $G_{IR}(f)$ (blue) and mean gains g_1, g_2 and g_3 for the frequency bands (red).

This processing was performed for every recorded logarithmic sweep resulting in frequency-band-dependent gain vector $\mathbf{g}_j(\varphi,\theta)$ holding the gains g_j for every sampling point of the enveloping surface ($j = 1, 2, 3$). Figures 13 (a), 14 (a) and 15 (a) are showing the 540 point LS radiation patterns for a Genelec 8020 LS for 100 Hz to 1 kHz, 1kHz to 5kHz and above 5 kHz.

For representation of the radiation patterns in a more compact way, the spherical harmonics (SH) domain was used. To find the optimal spherical wave spectra $\mathbf{\Psi}_{n,m,j}$, a optimization according to eq. 28 using the MATLAB-function `fminunc` was performed.

$$\begin{aligned}\min_{\mathbf{\Psi}_{n,m,j}} e &= \min_{\mathbf{\Psi}_{n,m,j}} \left(\sqrt{||\hat{\mathbf{g}}_j(\varphi,\theta) - \mathbf{g}_j(\varphi,\theta)||^2}\right) \\ &= \min_{\mathbf{\Psi}_{n,m,j}} \left(\sqrt{||(||\mathbf{Y}_{n,m}(\varphi,\theta)\mathbf{\Psi}_{n,m,j}||) - \mathbf{g}_j(\varphi,\theta)||^2}\right)\end{aligned} \tag{28}$$

Whereas $\mathbf{Y}_{n,m}(\varphi,\theta)$, $\mathbf{\Psi}_{n,m,j}$ and $\mathbf{g}_j(\varphi,\theta)$ have following structure:

$$\mathbf{Y}_{n,m}(\varphi,\theta) = \underbrace{\begin{pmatrix} \mathbf{y}_{n,m}(\varphi_1,\theta_1) \\ \mathbf{y}_{n,m}(\varphi_2,\theta_2) \\ \dots \\ \mathbf{y}_{n,m}(\varphi_N,\theta_N) \end{pmatrix}}_{[540\text{x}M]}, \quad \mathbf{\Psi}_{n,m,j} = \underbrace{\begin{pmatrix} \psi_1 \\ \psi_2 \\ \dots \\ \psi_M \end{pmatrix}}_{[M\text{x}1]}, \quad \mathbf{g}_j(\varphi,\theta) = \underbrace{\begin{pmatrix} g_j(\varphi_1,\theta_1) \\ g_j(\varphi_2,\theta_2) \\ \dots \\ g_j(\varphi_N,\theta_N) \end{pmatrix}}_{[540\text{x}1]}$$

$\mathbf{Y}_{n,m}(\varphi,\theta)$ holds the SH coefficients $\mathbf{y}_{n,m}(\varphi_k,\theta_k)$ for each measurement point at angle φ_k and θ_k. $\hat{\mathbf{g}}^i(\varphi,\theta)$ represents the via SH approximated gains. The radiation characteristic for a LS in the lower frequency area can be seen more likely ball-shaped and becomes more directed with increasing frequencies. Therefore, 1st-order SH are used for the directivity pattern for 100Hz to 1kHz, 3rd-order SH for 1kHz to 5kHz and 5th-order SH above 5kHz. With this simplification, the optimization process is more efficient by loosing less accuracy. Figures 13 (b), 14 (b) and 15 (b) are showing the approximated LS radiation patterns for a Genelec 8020 LS for 100 Hz to 1 kHz, 1kHz to 5kHz and above 5 kHz.

The equidistant microphone placement on the vertical ring of the setup together with the turning angles of the turntable results in a inhomogeneous distribution of the measurement points. So the radiation patterns from figures 13 to 15 (b) have a higher resolution at the poles but a lower resolution at the equator. To overcome this issue a homogeneous sampling of the spherical surface can by applied via a t-design [34, 36]. A t-design discretizes an arbitrary spherical polynomial $\mathcal{P}_n(\mu)$ with a limited degree ($n \leq t$) such that the discrete sum and the integral are equivalent.

$$\frac{4\pi}{L}\sum_{l=1}^{L}\mathcal{P}_n(\mu_l) = 2\pi\int_{-1}^{1}\mathcal{P}_n(\mu)d\mu \tag{29}$$

Whereas $\mu = \langle\boldsymbol{\theta}_S,\boldsymbol{\theta}\rangle$ is a continuous and $\mu = \langle\boldsymbol{\theta}_S,\boldsymbol{\theta}_l\rangle$ is a discrete variable. Direction $\boldsymbol{\theta} = (\varphi,\theta)$ is continous and $\boldsymbol{\theta}_l = (\varphi_l,\theta_l)$ are the discrete directions of the t-design. $\boldsymbol{\theta}_S = (\varphi_S,\theta_S)$ is the source panning direction. The t-designs play an important part at Ambisonic playback. Designs with $t \geq 2N+1$ representing optimal LS arrays where both energy $\mathbf{E}$ and energy vector $\mathbf{r}_E$ are panning-invariant. In this work, a 240-point 21-design from fig. 12 was applied only to have a more consistent visualization of the LS radiation patterns (cf. fig. 13 (c), 14 (c) and 15 (c)).

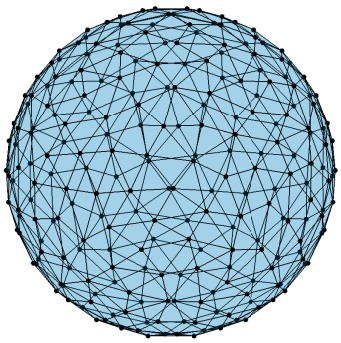


Figure 12 – Applied 21-design for homogeneous sampling of the SH radiation patterns.

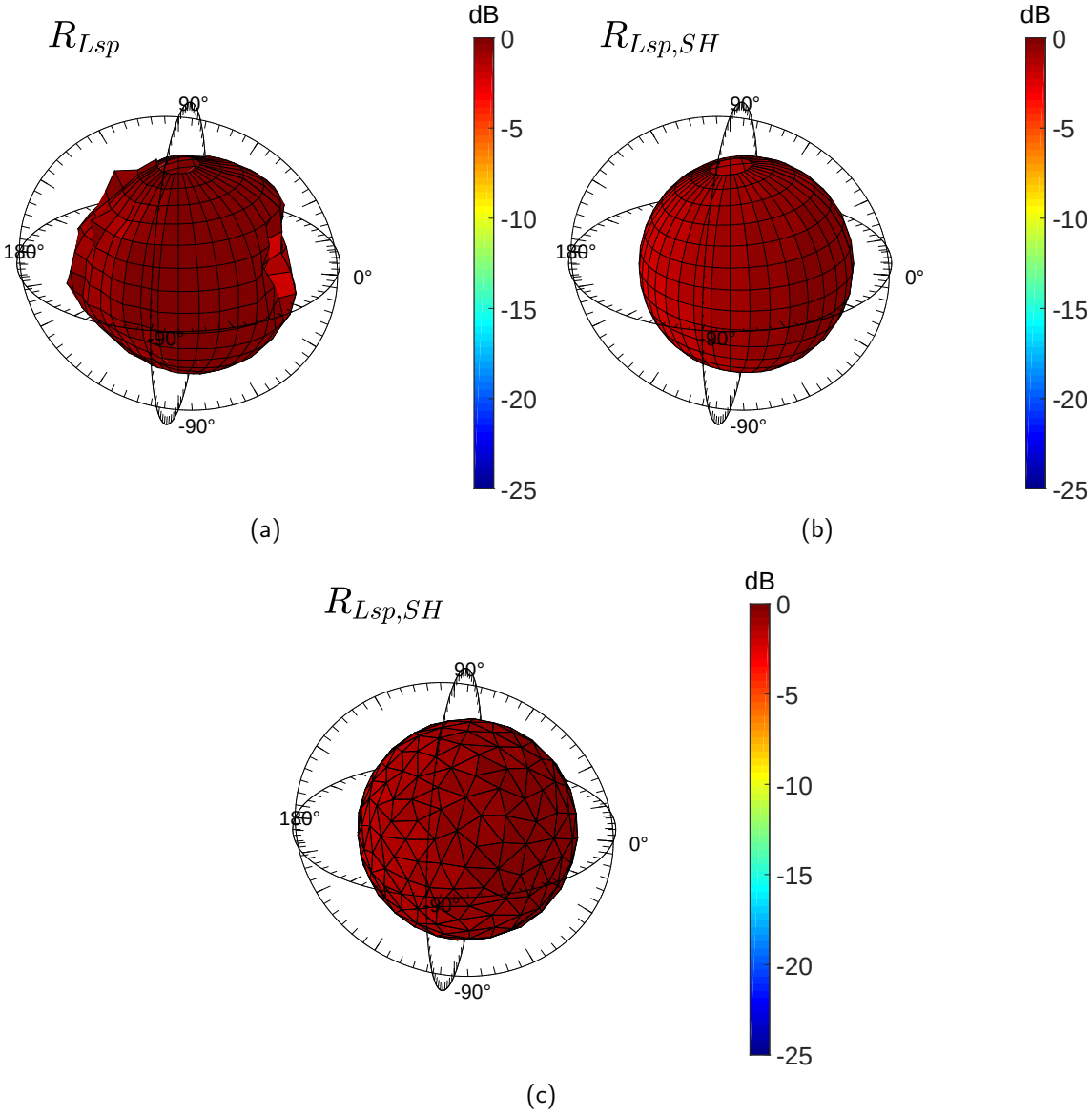


Figure 13 – Radiation patterns for a Genelec 8020 LS for the frequency band from 100 Hz to 1 kHz. (a): Radiation pattern from measurement. (b): Approximated radiation pattern in 1st-order spherical harmonics. (c): Radiation pattern in 1st-order spherical harmonics sampled on a 21-design.

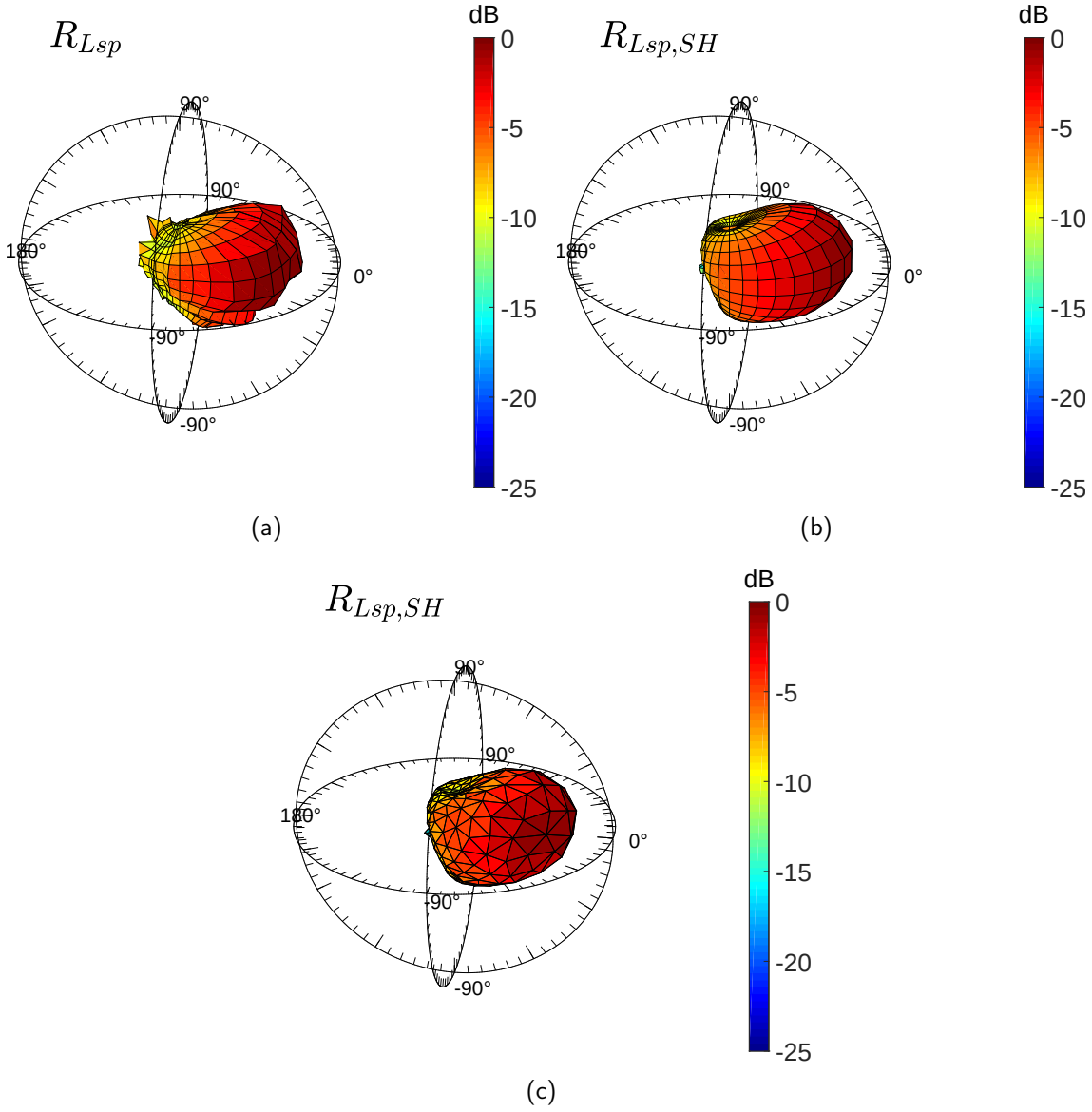


Figure 14 – Radiation patterns for a Genelec 8020 LS for the frequency band from 1 kHz to 5 kHz. (a): Radiation pattern from measurement. (b): Approximated radiation pattern in 3rd-order spherical harmonics. (c): Radiation pattern in 3rd-order spherical harmonics sampled on a 21-design.

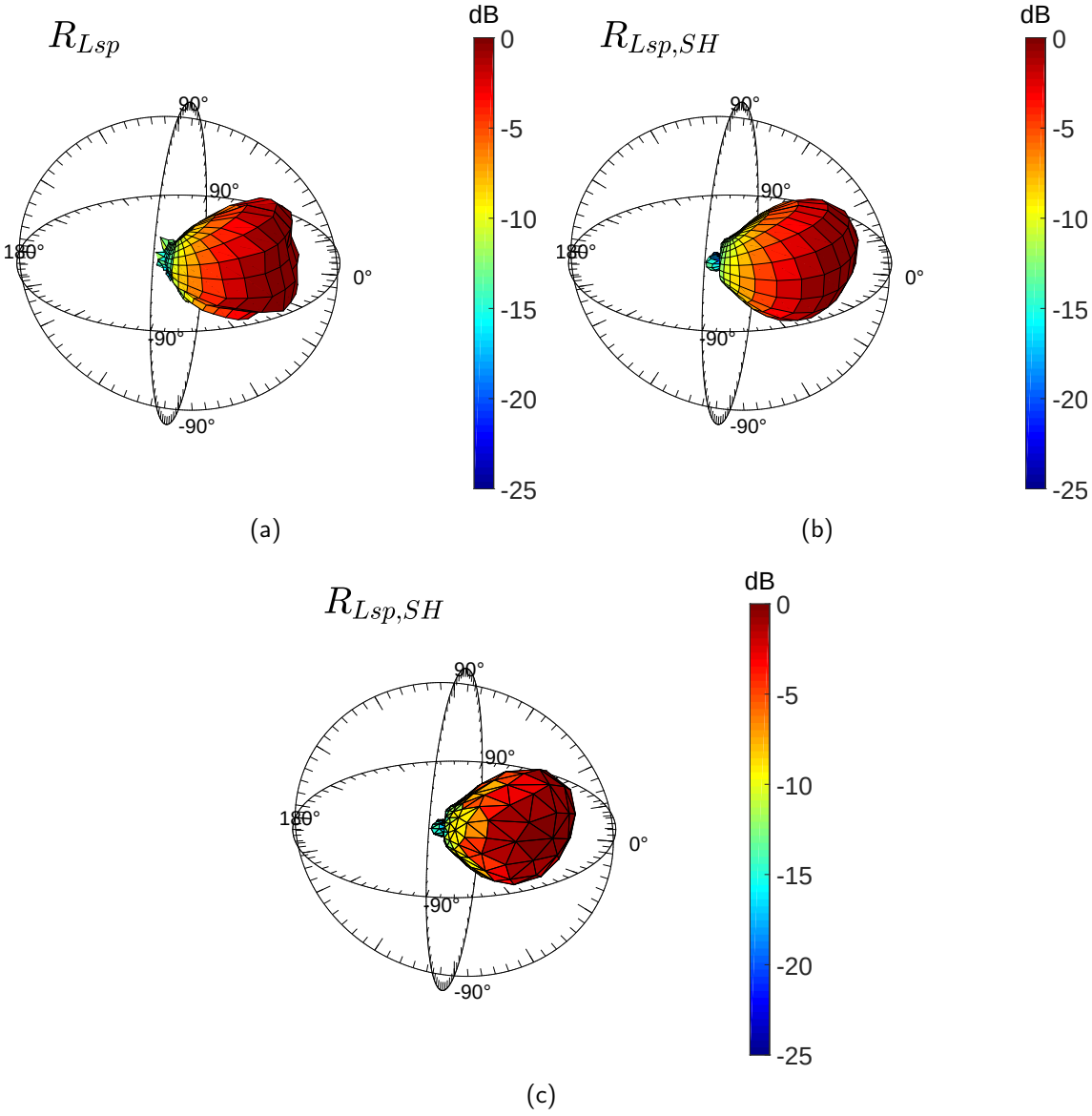


Figure 15 – Radiation patterns for a Genelec 8020 LS for the frequency band above 5 kHz. (a): Radiation pattern from measurement. (b): Approximated radiation pattern in 5th-order spherical harmonics. (c): Radiation pattern in 5th-order spherical harmonics sampled on a 21-design.

Moreover radiation patterns for a Lambda Labs CX1A, a Tannoy 1200, a Neumann KH120A and a Yamaha DXR8 LS were measured. All radiation patterns can be found in the appendix (see fig. 60 to 63). Measurement data as well as visualization and analysis tools can be found at the web page of the DirPat project [4] (see also in [5]).

In order to integrate the LS radiation patterns into the $\mathbf{r_E}$-vector model the patterns have to be applied on the LS positions with exact rotation. Off course this holds also for the mirrored LS positions. To illustrate this processing step, the radiation pattern for

4. `https://opendata.iem.at/projects/dirpat/`

a Genelec 8020 LS is applied to the LS positions of the listening experiment simulation environment.

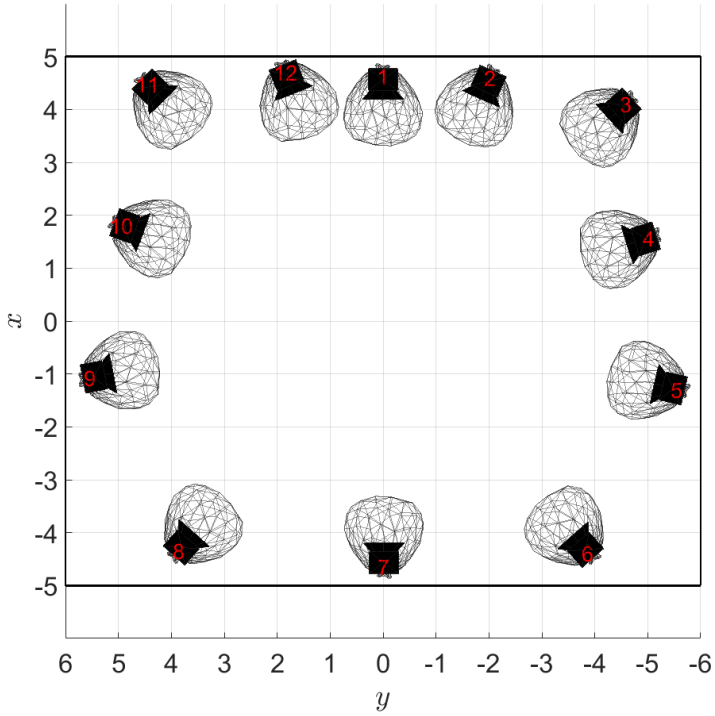


Figure 16 – Radiation patterns on LS positions.

The piercing angles $\hat{\varphi}_i$ and $\hat{\theta}_i$, on which the radiation pattern for the i-th LS has to be selected, can be determined with the source-to-receiver-vector $\mathbf{r}_{\boldsymbol{\theta}_i,\mathbf{v}}$ for the listening position $\mathbf{v}$.

$$\mathbf{r}_{\boldsymbol{\theta}_i,\mathbf{v}} = \mathbf{v} - \boldsymbol{\theta}_i = \begin{pmatrix} v[1] \\ v[2] \\ v[3] \end{pmatrix} - \begin{pmatrix} \theta_i[1] \\ \theta_i[2] \\ \theta_i[3] \end{pmatrix} = \begin{pmatrix} r_{\boldsymbol{\theta}_i,\mathbf{v}}[1] \\ r_{\boldsymbol{\theta}_i,\mathbf{v}}[2] \\ r_{\boldsymbol{\theta}_i,\mathbf{v}}[3] \end{pmatrix} \tag{30}$$

The transformation of a vector $\mathbf{p}$ from cartesian to spherical coordinates can be done in MATLAB with the function `cart2sph`, which implements the following equations:

$$\varphi = \tan^{-1}\left(\frac{p[2]}{p[1]}\right) \tag{31}$$

$$\theta = \tan^{-1}\left(\frac{p[3]}{\sqrt{p[1]^2 + p[2]^2}}\right) \tag{32}$$

$$r = \sqrt{p[1]^2 + p[2]^2 + p[3]^2} \tag{33}$$

Consequently the source-to-receiver-angles $\tilde{\varphi}_{\boldsymbol{\theta}_i,\mathbf{v}}$ and $\tilde{\theta}_{\boldsymbol{\theta}_i,\mathbf{v}}$ can be determined in the

following way:

$$\tilde{\varphi}_{\boldsymbol{\theta}_i,\mathbf{v}} = \tan^{-1}\left(\frac{r_{\boldsymbol{\theta}_i,\mathbf{v}}[2]}{r_{\boldsymbol{\theta}_i,\mathbf{v}}[1]}\right), \quad \tilde{\theta}_{\boldsymbol{\theta}_i,\mathbf{v}} = \tan^{-1}\left(\frac{r_{\boldsymbol{\theta}_i,\mathbf{v}}[3]}{\sqrt{r_{\boldsymbol{\theta}_i,\mathbf{v}}[1]^2 + r_{\boldsymbol{\theta}_i,\mathbf{v}}[2]^2}}\right). \tag{34}$$

To determine $\hat{\varphi}_i$ and $\hat{\theta}_i$, it is simply possible to subtract the LS rotation angles $\varphi_{LS,i}$ and $\theta_{LS,i}$ from $\tilde{\varphi}_{\boldsymbol{\theta}_i,\mathbf{v}}$ and $\tilde{\theta}_{\boldsymbol{\theta}_i,\mathbf{v}}$.

$$\hat{\varphi}_i = \tilde{\varphi}_{\boldsymbol{\theta}_i,\mathbf{v}} - \varphi_{LS,i} \quad \text{and} \quad \hat{\theta}_i = \tilde{\theta}_{\boldsymbol{\theta}_i,\mathbf{v}} - \theta_{LS,i}. \tag{35}$$

Now it is possible to sample the spherical harmonics $\mathbf{y}_{n,m}$ at a certain point $(\hat{\varphi}_i, \hat{\theta}_i)$ and multiply it with the spherical wave spectra $\boldsymbol{\Psi}_{n,m,j}$:

$$w_{LS,i,j} = \mathbf{y}_{n,m}(\hat{\varphi}_i, \hat{\theta}_i) \cdot \boldsymbol{\Psi}_{n,m,j}. \tag{36}$$

With eq. 36 we get the gain $w_{LS,i,j}$ at the point of the radiation pattern were it is pierced by the source-to-receiver-vector $\mathbf{r}_{\mathbf{u}_i,\mathbf{v}}$. This gain can be used to again extend the existing vector model from eq. 24:

$$\mathbf{r}_{\mathbf{E}} = \frac{\sum_{i=1}^{I}\left(|w_{LS,i,j}(\hat{\varphi}_i,\hat{\theta}_i)w_{\tau,i}w_{d,i}g_i|^{\overline{\gamma}_j}\boldsymbol{\theta}_i\right)}{\sum_{i=1}^{I}\left(|w_{LS,i,j}(\hat{\varphi}_i,\hat{\theta}_i)w_{\tau,i}w_{d,i}g_i|^{\overline{\gamma}_j} + \sum_{l=1}^{L}|(1-\overline{\alpha})^{K/2}w_{LS,i,l,j}(\hat{\varphi}_{i,l},\hat{\theta}_{i,l})w_{\tau,i,l}w_{d,i,l}g_i|^{\overline{\gamma}_j}\boldsymbol{\theta}_{i,l}\right)}{+\sum_{l=1}^{L}|(1-\overline{\alpha})^{K/2}w_{LS,i,l,j}(\hat{\varphi}_{i,l},\hat{\theta}_{i,l}))w_{\tau,i,l}w_{d,i,l}g_i|^{\overline{\gamma}_j}\Big)}. \tag{37}$$

Whereas the term $|w_{LS,i,j}(\hat{\varphi}_i,\hat{\theta}_i)w_{\tau,i}w_{d,i}g_i|^{\overline{\gamma}_j}\boldsymbol{\theta}_i$ stands for the gain vector of the i-th LS and $\sum_{l=1}^{L}|(1-\overline{\alpha})^{K/2}w_{LS,i,l,j}(\hat{\varphi}_i,\hat{\theta}_i)w_{\tau,i,l}w_{d,i,l}g_i|^{\overline{\gamma}_j}\boldsymbol{\theta}_{i,l}$ represents the gain vectors of its mirrored versions.

Figures 17 (a) to (d) are showing the effect of the directed radiation patterns from fig. 14 and 15 on the simulation with the image-source model for the mean localization error $\overline{\Delta\theta}$. Playback conditions, $\hat{T}_{60}$ resp. $\overline{\alpha}$ are similar to those from fig. 8. The frequency-dependent panning with a slope of $\overline{\gamma}_2 = 1.637$ resp. $\overline{\gamma}_3 = 1.967$ is used for this simulation according the chosen radiation pattern (1kHz $\leq f \leq$ 5kHz resp. $f >$ 5kHz). In fig. 17 (a) and (b) a 1st-order image-source model is deployed. Comparing them with fig. 8 (a) shows the influence of a more directed source very good. With higher radiation directivity the room is excited less and therefore disturbing reflections disappear. Especially for a high directivity (fig. 17 (b)) it is clearly visible that the area with good localization inside the array becomes slightly larger and outside plausible localization becomes possible again. At fig. 17 (c) and (d) the 2nd-order model is used. Again a higher directivity in radiation (fig. 17 (d)) leads to better results for localization. Interestingly localization

close to the room boundary surfaces is worse because of the minimal distance between the LSs and the boundary surfaces. Nevertheless in the room corners, it becomes better.

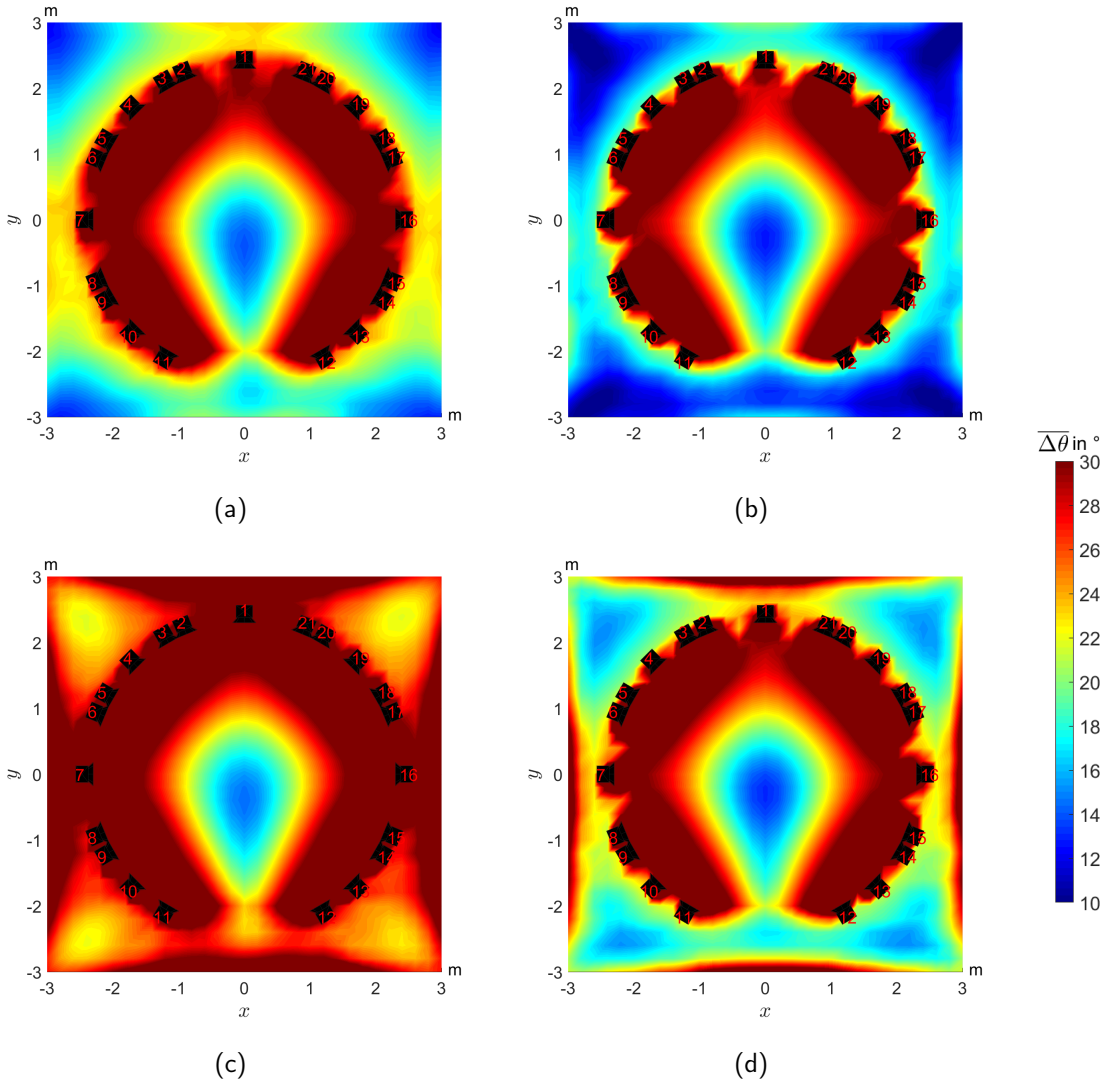


Figure 17 – Mean Localization error $\overline{\Delta\theta}$ in dependence on the listening position for 5th-order Ambisonics and max-$\mathbf{r_E}$ weighting with applied 3-D image-source model and radiation directivity of the LSs. No gain and delay compensation is done. $\overline{\alpha} = 0.35$ is computed with eq. 18 and a $\hat{T}_{60} = 0.224$s. Genelec 8020 radiation patterns are used for simulation. A 1st-order ((a), (b)) and a 2nd-order ((c), (d)) image-source model is simulated. (a), (c): Radiation pattern from fig. 14 is used. (b), (d): Radiation pattern from fig. 15 is used.

3 Spatial hearing with two sound sources

The simplest kind of spatial hearing with multiple sound sources is the stereophonic case. For investigations in this domain, it is in general assumed that the two sound sources, resp. LSs, are playing exactly the same signal. Playback situation only can be influenced by ICLDs and/or ICTDs. If this differences are deployed to the stereophonic setup there are in general three different cases for the direction of the auditory event:

1. The direction of the auditory event is dependent on the direction of the two LSs and its reproduced signals:
   This case is generally called **summing localization** (cf. [73]). It occurs for a small ICLD ($\Delta L < 18$dB) and ICTD ($\Delta t < 1$ms). Here, the ear interprets the two incoming signals as one single auditory event. A so-called phantom sound source is perceived. Some listening experiments regarding this effect can be found in subsec. 3.1.
2. The direction of the auditory event is dependent on the direction of **one** LS and its reproduced signal:
   If a ICLD of $\Delta L \geq 18$dB resp. ICTD of $\Delta t \geq 1$ms occurs the auditory event is perceived at the place of the louder resp. leading sound source. Especially the ICTD is an important subject in previous investigations (cf. [18, 45]). Cremer formulated the „**Gesetz der ersten Wellenfront**“ in 1948 (cf. [9]) which is also well known under the so-called „**Precedence-effect**“ [71]. It states that only the direction of the wave front that is first arriving at the ear is mainly important for localization. The other lagged arriving signal is masked by this first wave front up to a certain delay. For further explanations see subsec. 3.2.
3. There are two auditory events which depend each on the direction of the respective LS and its reproduced signal:
   If the delay between the leading and the lagging signal reaches a certain amount, the single auditory event splits into a sequence of two auditory events. The lagging signal is called **echo**. The amount of delay for this effect, called „echo threshold“, depends on various parameters such as the spectrum or the envelope of the signal (more details see subsec. 3.3).

3.1 Summing localization

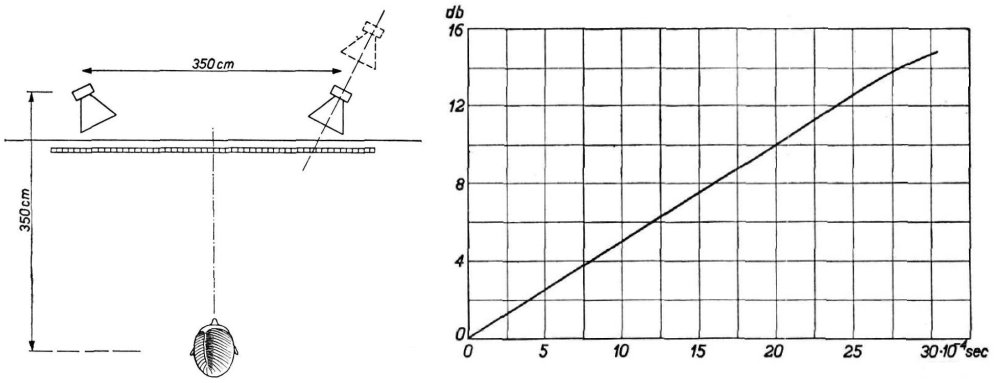
Many listening experiments about the correlation between ICLDs and ICTDs regarding summing localization has been done in previous investigations (e.g. in [14, 18, 44, 45]). The results of all these listening test can be represented in curves, which relate the ICTD Δt (in ms) to a certain ICLD ΔL (in dB). Basically the listening experiment and consequently the resulting curves can be divided in two groups.

- Curves regarding equivalence stereophony:
  This curves specify how much the ICLD ΔL displaces the phantom sound source out of the center position in the same way as an equivalent ICTD Δt.

- Trading curves:
  This curves specify how much ICLD ΔL is needed to compensate a certain amount of ICTD Δt to hold the phantom sound source at the center position.

Off course, this two different approaches are leading to different results for the ratio $\frac{\Delta L}{\Delta t}$ (in $\frac{\text{dB}}{\text{ms}}$). In the following, some curves from chronologically ascending investigations are depicted and discussed.

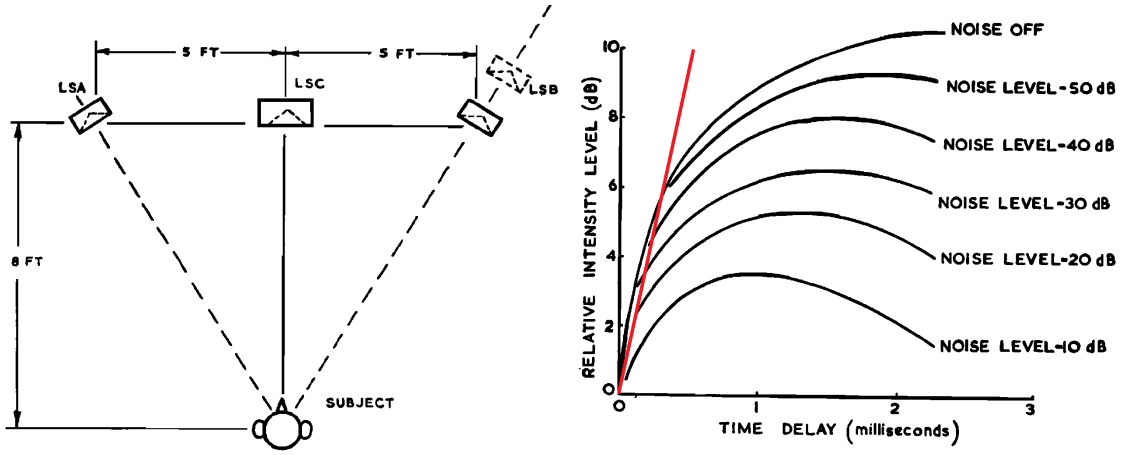
Figure 18 (b) shows the trading curve which was measured in the listening experiments by de Boer during his dissertation in 1940 [14]. Measurement was done with the setup that is depicted in fig. 18 (a). To produce an ICTD, the right LS of the setup was displaced to the back and its level was adjusted for equal loudness from left and right LS at the listening position. As stimulus a speech recording was played back and the subjects were asked to adjust the level of the right LS to perceive the phantom sound source from the center position. The subject could turn his head slightly during the experiment but were asked to face to the center position. De Boer found a direct proportional relation $\overline{\frac{\Delta L}{\Delta t}} \approx 18.3\ \frac{\text{dB}}{\text{ms}}$.



(a) (b)

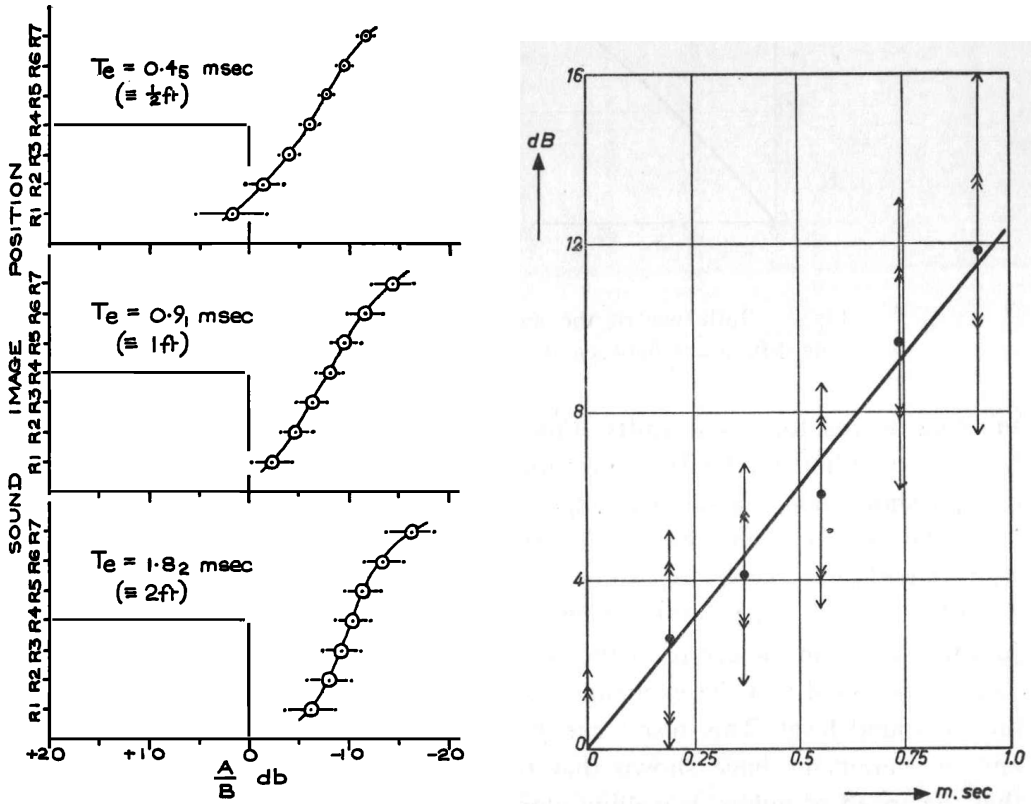
Figure 18 – Schematic setup (a) and trading curve (b) from investigations of de Boer in 1940 [14]. Mean interchannel time-level-ratio $\frac{\Delta L}{\Delta t} \approx 18.3\ \frac{\text{dB}}{\text{ms}}$.

Leakey measured some trading curves under different ambient noise levels in 1957 [44]. The setup of his listening experiment is shown schematically in fig. 19. The listening experiment was done in a similar manner as those from de Boer 1940. As reference an additional LS was placed in the center position of the stereophonic panorama and the subject should adjust the playback direction of the phantom sound source to this direction. The stimuli included single component tones as well as random noise. In this listening experiment a headrest was installed to minimize head movements. Leakey found a non-linear relation between ΔL and Δt. If linearisation is done for the case without noise (red line) the mean interchannel time-level-ratio is $\overline{\frac{\Delta L}{\Delta t}} \approx 18\ \frac{\text{dB}}{\text{ms}}$.



(a) (b)

Figure 19 – Schematic setup (a) and trading curve (b) from investigations of Leakey in 1957 [44]. Mean interchannel time-level-ratio $\overline{\frac{\Delta L}{\Delta t}} \approx 18\ \frac{\text{dB}}{\text{ms}}$.



(a) (b)

Figure 20 – (a): Displacement of panning curve in dependence on the ICTD Δt from investigations of Leakey in 1959 [45] ($\overline{\frac{\Delta L}{\Delta t}} \approx 8.9\ \frac{\text{dB}}{\text{ms}}$). (b): Trading curve from investigations of Franssen in 1961 [18]. Mean interchannel time-level-ratio $\overline{\frac{\Delta L}{\Delta t}} \approx 12.4\ \frac{\text{dB}}{\text{ms}}$.

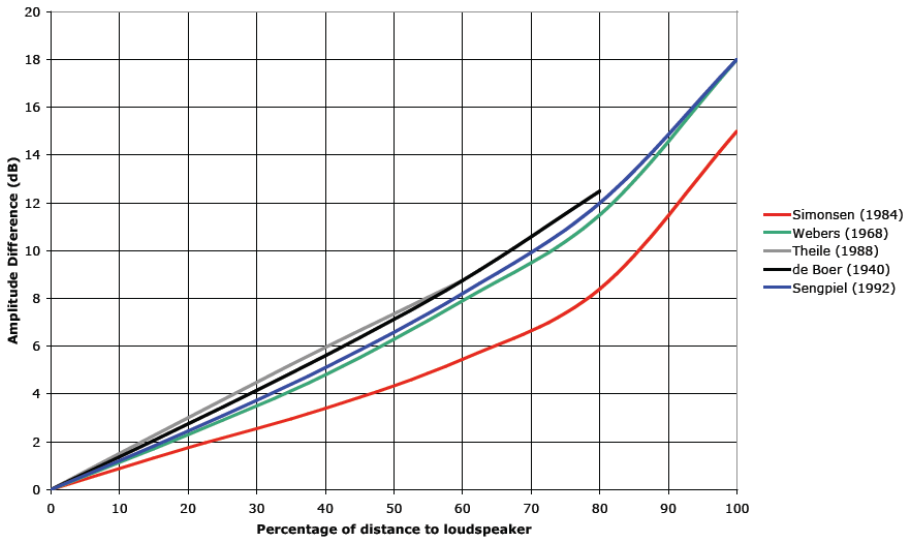
Another investigation about the relation between ΔL and Δt was done by Leakey in 1959 [45]. This investigation was done with the same listening experiment setup as in 1957. Figure 20 (a) shows the displacement of the panning curve in dependence on the ICTD Δt. From this displacements, it is possible to determine a $\overline{\frac{\Delta L}{\Delta t}} \approx 8.9$ $\frac{\text{dB}}{\text{ms}}$. Also Franssen did some research to the topic in his dissertation 1961, but with a listening experiment via headphones (cf. [18]). Therefore signals of the left and right ear were totally independent from each other and crosstalk was not possible. A trading experiment was performed and he found linear a relation of $\overline{\frac{\Delta L}{\Delta t}} \approx 12.4$ $\frac{\text{dB}}{\text{ms}}$ (cf. fig. 20 (b)).

The curves for the ICLD and ICTD that are depicted in fig. 21 were combined by Martin in 2006 [52]. For our investigations, only the curves of Simonsen (red) and Sengpiel (blue) are of interest. Simonsen performed a listening test with a stereo LS setup in 1984 [67]. Clicks from maracas and claves were used as stimuli signals. The listening test from Sengpiel was performed in a studio room setup in 1992. He used monaural human speech as well as classical music as stimuli signals. It is clearly visible that Simonsen gets a lower slope for his panning curves than Sengpiel. Furthermore there seems to be a higher slope for more deflected sound sources in both curves for ICLD and ICTD. This behaviour is more distinct in the results from Simonsen. Also Lee and Rumsey made this observation in there investigations 2013 [46]. Based on the curves of fig. 21 it is possible to compute the ratio $\overline{\frac{\Delta L}{\Delta t}}$. Table 2 is showing the computed values for the data of Simonsen and Sengpiel as well as from the other discussed curves.

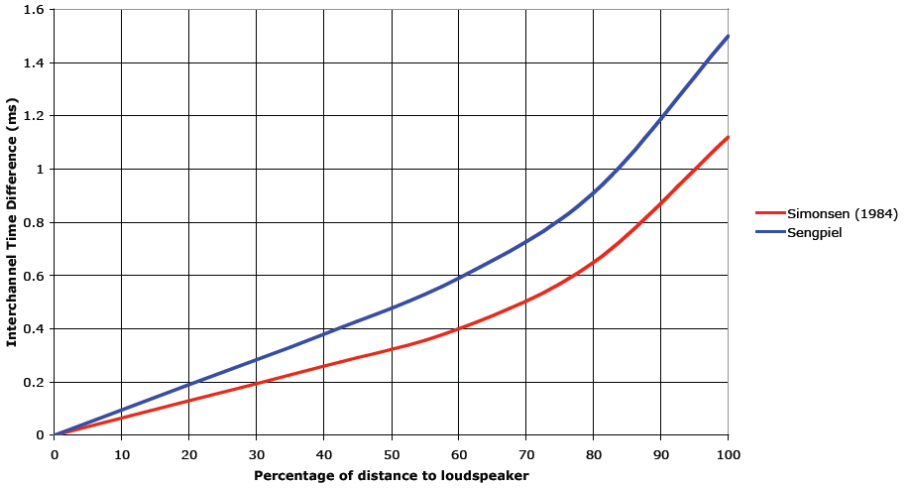
Type	**List. exp.**	$\overline{\frac{\Delta L}{\Delta t}}$ **in** $\frac{\text{dB}}{\text{ms}}$
Equiv. stereophony	Leakey (1959)	8.9
	Simonsen (1984)	13.6
	Sengpiel (1992)	13.3
Trading	de Boer (1940)	18.3
	Leakey (1957)	18
	Franssen (1961)	12.4

Table 2 – Mean interchannel time-level-ratios $\overline{\frac{\Delta L}{\Delta t}}$.

Comparing the values from table 2 for the two types of listening experiments shows that a $\overline{\frac{\Delta L}{\Delta t}}$ resulting from trading tends to be higher than from equivalence stereophony. It seems to be that on average a $\overline{\frac{\Delta L}{\Delta t}} \approx 14$ $\frac{\text{dB}}{\text{ms}}$ is reasonable. Of course this value is dependent on the aperture and size of the stereo setup, spectrum and temporal envelop of the presented stimulus, the experimental room, the used loud speakers and others. Moreover the permission for head movements during the listening experiment is a big influence on the resulting $\overline{\frac{\Delta L}{\Delta t}}$. All investigations that are presented in this subsection assume a frontal playback situation. It is questionable if this results can be generalized to lateral playback situations.



(a)



(b)

Figure 21 – (a): ICLD panning curves. (b): ICTD panning curves. Combined by Martin in 2006 [52].

A good qualitative depiction of the connection between ICLD and ICTD for stereophonic playback was done by Blauert in [4] (see fig. 22). In this figure, consensual as well as opposing combination of both panning methods is visible in a quite simple manner. Consensual combination of level and time difference results in a more deflected phantom sound source. By contrast opposing combination results in a less deflected phantom sound source.

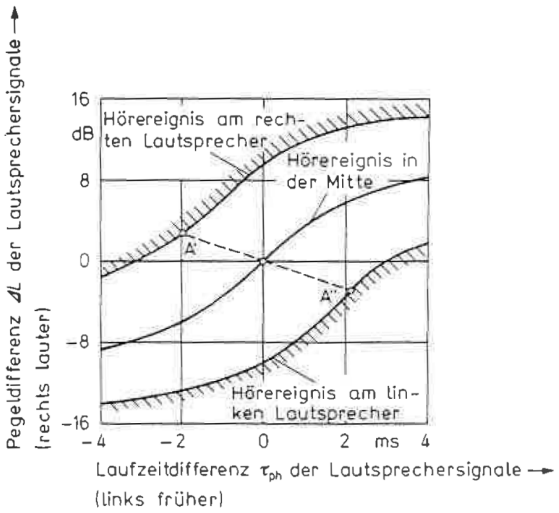


Figure 22 – Direction of the phantom sound source depending on ICLD and ICTD [4]. According to work from Franssen [19].

3.2 Precedence effect

As described in the introduction of the section 3 the precedence effect is a very important aspect of spatial hearing. It helps the human to distinguish the true direction of a sound source even when there are other corrupting cues. This cues could be room reflections or other sound sources emitting coherent signals which do not arrive simltaniously at the listening position. Therefore, the effect is to consider when studying localisation outside the center position of a spherical LS array. Literature states that precedence is of importance when Δt becomes greater than a certain threshold of 0.63ms $< \Delta t_p <$ 1ms [4]. Furthermore there is an upper threshold for Δt at which the single acoustical event splits up into two events and an echo is perceptible. This so called echo threshold is very hard to define, because it is dependent on various conditions (see subsec. 3.3). The precedence effect is still valid even if the level between leading and lagging signal is $\Delta L =$ -10dB [35]. This behaviour was discovered by Haas and holds for a $\Delta t \leq$ 30ms.

Most previous investigations regarding the precedence effect were carried out with the help of lead-lag-experiments. These experiments can be done by playing back a stimulus over a stereo LS setup whereas playback is delayed on one LS. A common used stimulus signal is a short click so that for most delay lengths, a small or no time overlapping happens for leading and lagging signal. Also a lead-lag-experiment can be performed via headphones. In this case, each of the two signals is played back with a certain synthetic interaural time difference to both sides of the headphone. In this way, playback of the leading and lagging stimulus can be modelled.

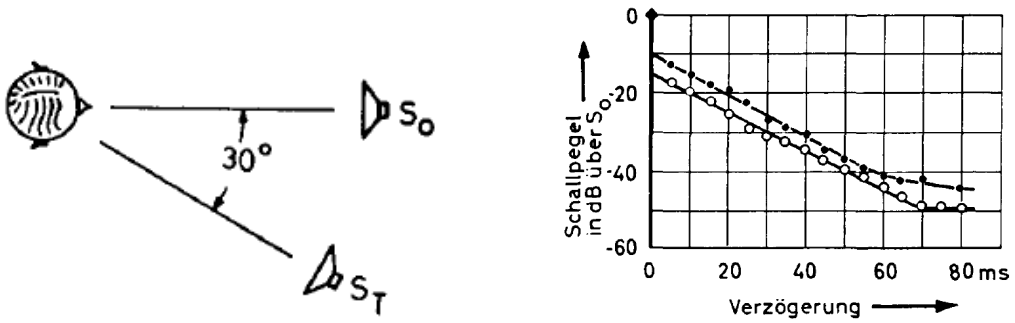
The temporal envelope and especially the onset of the source signal plays an important role in the precedence effect [60]. The effect is less dominant for slow ongoing resp. more stationary signals then for transient ones. Also the spectrum of the source signal is an important influence. Moreover it could be found that the effect is weaker when

leading and lagging signal both have the same interaural time difference. [66, 71]. This is the case when both sound sources are placed in lateral direction to the listener. The investigation from [59] states that sound sources on the same cone-of-confusion do not produce any precedence.

Stitt tried to integrate the precedence effect as well as the behaviour at the cone-of-confusion in his extension for the $\mathbf{r_E}$-vector model (cf. [69]). It could be shown in [43] that the computational effort from Stitt's extension do not result in better localization prediction in comparison to a simpler $\mathbf{r_E}$-extension (cf. subsec. 2.2).

3.3 Echo and echo threshold

As implied in subsec. 3.2, the threshold for perception of an echo is not clearly defined because of influences from various changes of an auditory event. The transition from an occurring single sound source to an echo characterized by changes like timbre, source width or blurriness of the phantom sound source. Therefore, there are several definitions for the echo detection threshold. The lowest possible threshold value is the level of the lagging signal that causes minimal timbre changes of the primary/leading signal. This threshold was defined by Seraphim 1961 and is called „absolute threshold of perceptibility aWs " [65]. It could be also called just noticeable difference for echo perception. Figure 23 (b) shows the aWs measured by Seraphim with an assessment procedure (•) and a constancy method (○). This curve is the result of a listening test with the LS setup showed in fig. 23 (a) that played back a speech signal. Seraphim investigated the maximal level difference $\Delta S = S_T - S_0$ that is needed to do not hear the echo signal S_T with delay Δt. He obtains a linear ratio of $\tau = \frac{\Delta S}{\Delta t}$ = -0.5 $\frac{\text{dB}}{\text{ms}}$.



(a) (b)

Figure 23 – (a): Listening experiment setup from [65]. (b): Absolute threshold of perceptibility aWs measured with an assessment procedure (•) and a constancy method (○).

Another possible way to determine a threshold for echo perception is it to ask at which level of the lagging signal this signal is barely not perceived as an echo. Based on this definition we can talk about a real threshold for perception of an echo, also called echo

threshold. This threshold was measured in various previous investigations [6, 35, 53]. When it comes to a listening test design we have to take care about the playback situation. Playback with LSs and headphones are of course possible but will lead to different results.

In general the echo threshold is dependent on (according to [4]):

- (I) The impulsiveness of the leading resp. lagging signal:
  $\hookrightarrow$ More impulsive signals results in a lower echo threshold.
- (II) The level of leading and lagging signal:
  $\hookrightarrow$ More level of the signal results in a lower echo threshold.
- (III) The direction of arrival of the lagging signal in relation to the leading signal:
  $\hookrightarrow$ The more the lagging signal is arriving from another direction the lower is the echo threshold (also cf. [12]).
- (IV) The spectral characteristics of the lagging signal in comparison to the leading signal:
  $\hookrightarrow$ A lagging signal with boosted higher frequency components results in a lower echo threshold.
- (V) The impulse density of the leading signal:
  $\hookrightarrow$ Dense leading signals resulting in a higher echo threshold.
- (VI) Additional signals which mask leading and lagging signals:
  $\hookrightarrow$ More additional signals increase the echo threshold. Higher levels of this signals also lift the threshold.

Consequently when measuring the echo threshold via a listening experiment, the threshold criterion is a very critical aspect [12]. It has to be communicated very precisely to the subjects. Therefore experienced subjects are more suitable for those tests. Figure 24 shows the results from two listening tests [50, 53] with different threshold criteria. Lochner and Burger asked for the minimum level difference $\Delta L = L_{ST} - L_{S0}$ for a certain delay of the lagging signal S_T so that an echo is clearly audible. By contrast Meyer and Schodder asked for a ΔL so that an echo is just inaudible. Consequently, the curve from Meyer and Schodder tends to a smaller ΔL for the same time delay of S_T. Moreover the behaviour of the curve around $\Delta t = 20$ ms is also an effect of the used threshold criterion. For subjects of this listening experiment, the echo was still audible not just as an additional auditory event but also as a shift of the direction of the primary auditory event. Therefore this curve is to treat with caution. The curves from Lochner and Burger [50] are more suitable for our investigations. They are decreasing in a more consistent way with a slope of $\tau = \frac{\Delta L}{\Delta t} = \approx$ -0.33 $\frac{\text{dB}}{\text{ms}}$ for $L_{S0} = 50$ dB, resp. $\tau = \frac{\Delta L}{\Delta t} \approx$ -0.19 $\frac{\text{dB}}{\text{ms}}$ for $L_{S0} = 25$ dB. Moreover the dependence of the echo threshold from the level of the primary sound source can be seen.

Figure 25 shows results from investigations in [10]. It confirms the statement that more impulsive signals lead to a lower echo threshold. Here, noise impulses with a shorter duration lead to a steeper descending curve for $\frac{\Delta S}{\Delta t}$. If we perform a linear fit for the three curves for a delay time of 20 ms $\leq \tau_{ph} \leq$ 60 ms we get a slope of $\tau = \frac{\Delta L}{\Delta t} = \approx$ -0.84 $\frac{\text{dB}}{\text{ms}}$ for 10 ms, $\tau = \frac{\Delta L}{\Delta t} \approx$ -0.80 $\frac{\text{dB}}{\text{ms}}$ for 30 ms and $\tau = \frac{\Delta L}{\Delta t} \approx$ -0.70 $\frac{\text{dB}}{\text{ms}}$ for 100 ms noise impulse duration.

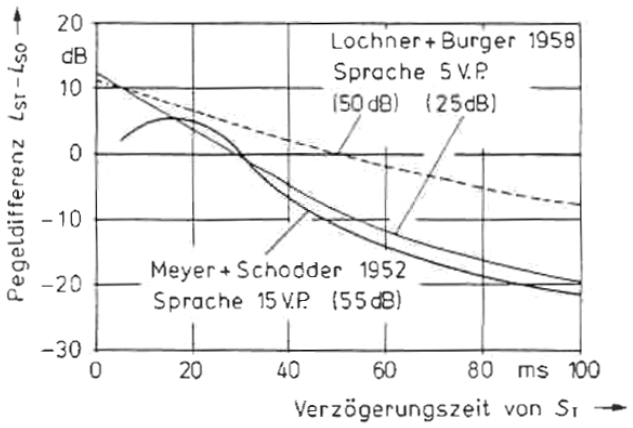


Figure 24 – Echo thresholds for stereo playback (aperture $\alpha = 40°$) of investigations from [50, 53]. The stimulus signal was speech with a mean speaking rate of $\approx$ 5 syllables/s.

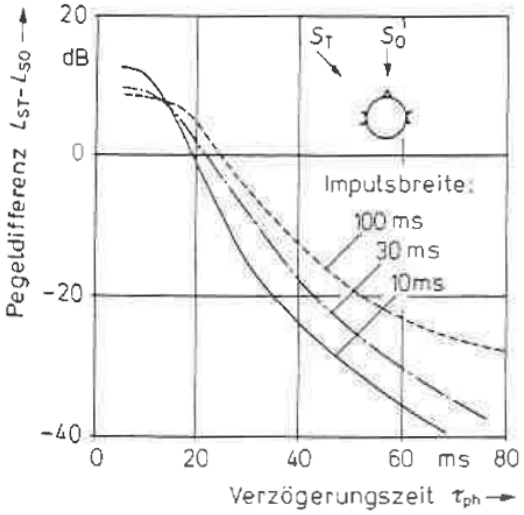


Figure 25 – Echo thresholds for noise impulses of different duration (10, 30 and 100 ms) [10]. The level of the primary auditory event is not known. S_0 front the front, S_T from the left/right.

Also Rakerd did some investigations regarding the echo threshold in 2000 [61]. As threshold criterion for the listening test, he asked for the minimum level difference ΔL at which a delayed copy of speech signal is barely audible as a distinct auditory event (exp. 1). Moreover he asked for the minimum level difference ΔL at which all audible effects caused by the lagging signal disappear (exp. 2). He did the listening test for the horizontal and the sagittal plane of five subjects (S1 to S5). Results from this listening experiment are shown in fig. 26. On this data the slope for auditory events in the horizontal plane of τ = -0.25 $\frac{\text{dB}}{\text{ms}}$ and in the sagittal plane of τ = -0.23 $\frac{\text{dB}}{\text{ms}}$ is obtained.

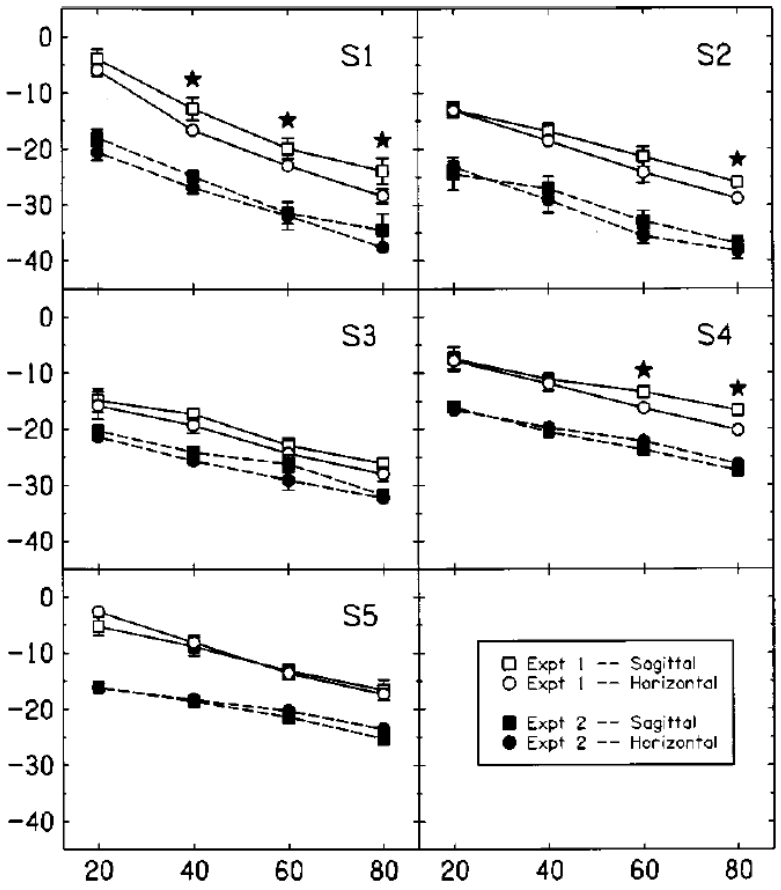


Figure 26 – Results from the lead-lag-experiment performed by Rakerd in 2000 [61]. Open symbols show echo thresholds for exp. 1. Filled symbols show masked thresholds for exp. 2.

Another investigation was done by Davis and Lee in 2016 [12]. They tried to find an upper and lower limit for the echo threshold by asking for two different threshold criteria for the same playback situation.

Criteria were:

- (1): The maximum ICTD where one single virtual sound source is barely perceptible,
- (2): The minimum ICTD where two separate sound sources are barely perceptible.

They found that the echo threshold for playback of pink noise is situated within a ICTD of approx. 50 ms to 10 ms depending mainly on the aperture angle of the LS pair. For playback of speech the interval for the ICTD is approx. 50 ms to 20 ms. Moreover results for horizontal and azimuthal aperture angles are located in the same value range for playback of pink noise and speech.

4 Listening experiment

For validation of the extensions of the $\mathbf{r_E}$-vector, respectively for tuning of the slope paramater τ, a listening experiment was performed. In this listening experiment the subjects had to adjust the playback direction of a moveable stimulus on a circle of LSs. The adjustment had to be done so that playback direction of the stimulus matched with this of another two copies of the stimulus. This two copies were played back at the same fixed direction of the LS circle. Playback direction only was adjusted via the ICTD between two appropriate selected LSs out of the LS circle. ICLD of the LS pair was set to 0 dB.

4.1 Conditions

Two different stimulus signals, pink noise and a click, were used for the listening experiment. The click was produced within a Pd patch (Appendix: Figure 59) with a unit sample sequence witch was low-pass-filtered at $f_{lp} = 1$kHz. The signals were presented as a sequence of three pink noise bursts respectively clicks following each other. This sequence was repeated in a loop until the subject confirmed its input. Figure 27 illustrates the sequence of the stimuli in time as well as the envelopes of the pulsed pink noise (PPN). The first two bursts/clicks of a sequence were played back from a fixed direction during one trial of the listening experiment. The third burst/click could be moved on the LS circle with the help of the shuttle/jog wheel via VBAP [56].

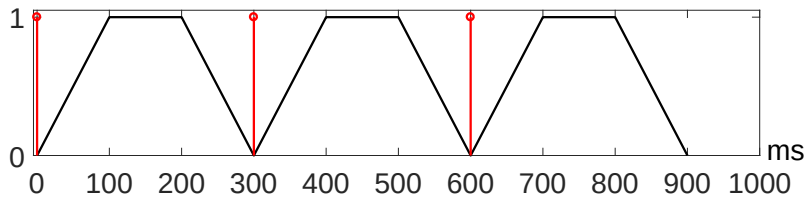


Figure 27 – Sequence of the clicks (red) in time as well as the envelopes of the PPN (black).

Every single envelope of the PPN consists of a attack, a sustain and a release section with 100ms each. Attack and release are chosen with this length to provide quasi stationary sound field conditions at playback. Conversely the click presents a more transient signal and generates non-stationary sound field conditions.

The following ICTDs were used to create phantom sources between the different LS pairs: 0ms, ±5ms, ±10ms, ±20ms and ±30ms. Figure 28 illustrates them for the first click of the click sequence.

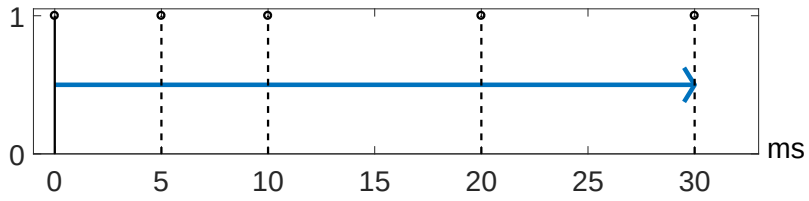


Figure 28 – Signal (click) for the leading channel (solid stem) and possible delayed versions for the lagging channel (dashed stems).

A positive sign in front of a time difference indicates a delay for the right hand LS related to the center direction (LS (1)) and the ±90° directions (LS (7) and (16)). The stimuli sequences were presented over the LSs listed in tab. 3 for even and odd numbered subjects. This constraint had to be done to reduce the total amount of presented stimuli and in further consequence the test duration per subject.

Subject	LS pairs		
	center	left	right
even	(1) (2,21) (3,20)	(7) (8,6) (10,4)	(16) (18,14) (20,12)
odd	(4,19) (5,18) (6,17)	(7) (9,5) (11,3)	(16) (17,15) (19,13)

Table 3 – LS pairs used for playback for even and odd numbered subjects.

For playback with a single LS ((1), (7) and (16)) there were obviously no time differences. Consequently each subject had to evaluate 2 (stimuli signals) × [3 (single LS) + 9 (level differences) × [5 (center LS pairs) + 4 (side LS pairs)]] = 168 stimuli sequences.

4.2 Setup

A circle of 21 Genelec 8020A LSs[5] with a radius of $r = 2.5$m and a height of $h = 1.2$m was built up around the center listening position of the IEM CUBE [78]. The IEM CUBE is a 11m × 11m × 5m experimental studio with a reverberation time within the limits of ITU-R BS.1116-1 [38]. The listening position was placed in the center position of the LS circle. Figure 29 shows the listening experiment setup schematically and tab. 4 lists the LS pairs according to their aperture angle α. Each LS was adjusted to an equal level of 60dB(A) at the listening position when playing broadband pink noise.

5. https://www.genelec.com/support-technology/previous-models/8020a-studio-monitor

α	45°	60°	90°	120°	135°
center LS pair	[2,21]	[3,20]	[4,19]	[5,18]	[6,17]
lateral LS pair	[6,8] [15,17]	[5,9] [14,18]	[4,10] [13,19]	[3,11] [12,20]	- -

Table 4 – Aperture angle α of all LS pairs.

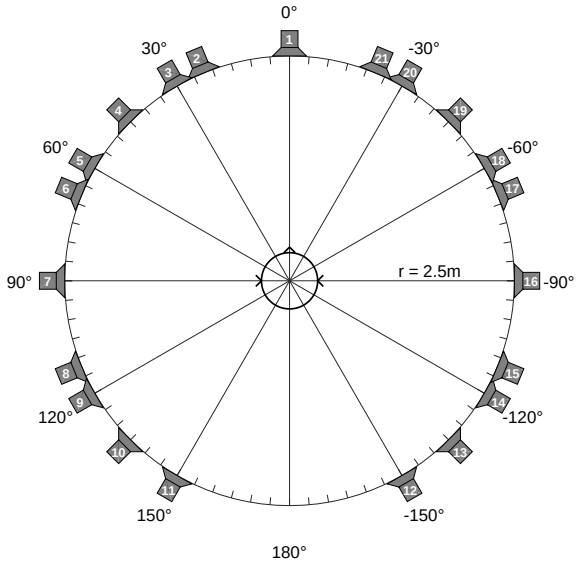


Figure 29 – Schematic setup of the listening experiment.

The LSs were connected to two 16 channel digital/analog converters M-16 DA [6] which were controlled via MADI. The MADI signal was sent by a MADIface USB [7] which was connected to a PC via USB. This PC ran the control software. For controlling of the listening experiment a patch (Appendix: Figure 59) was written in the open source software pure data [8] [58]. Figure 30 shows the flow diagram for the listening experiment.

6. `http://www.rme-audio.de/products/m32da.php`
7. `http://www.rme-audio.de/products/madiface_usb.php`
8. available on `http://puredata.info/downloads`

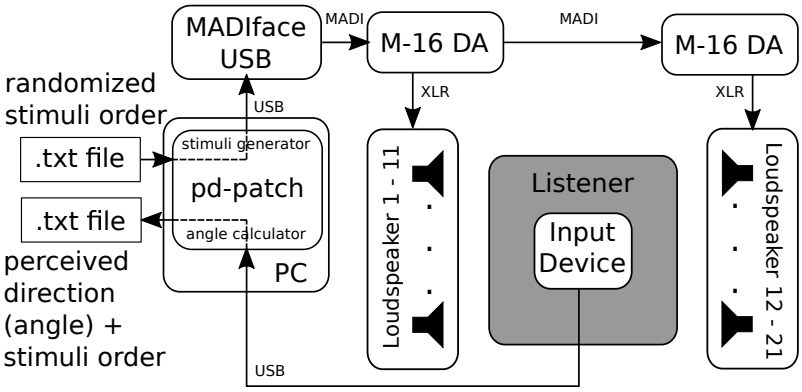


Figure 30 – Flow diagram of the listening experiment setup with all devices.

For one trial run, Pd read the current line in a text file that includes the randomized stimuli series with its parameters. The parameters in the line of the series were employed to the patch to generate the stimuli in real time according to figure 31. Trigger pulses with a interval duration of 300ms in a 1s loop were produced with a metronome and delay objects. The first two trigger pulses, which had a delay of 0ms and 300ms, were used to control playback of the two stimuli with fixed direction. They triggered the starting point of the envelopes respectively the dirac-delta impulses. The envelopes were multiplied with the pink noise generator to produce the PPN. The dirac-delta impulses were filtered at 1kHz with a one-pole low pass filter to produce the clicks. The playback direction of the first two bursts/clicks is produced only by ICTDs of the LSs chosen in the "2 out of 21 switch" according to Table 3. If there was playback only for one LS ([1], [7] and [16]) the "right" channel, related the main playback directions 0° and ±90°, was muted. The trigger pulse for the pannable stimulus had a delay of 600ms. The noise burst/click was produced analogical two the first two fixed direction stimuli. The VBAP block was a simple lookup table, which held the precomputed gains for all 21 LSs for every panning direction with a resolution of 1° steps.

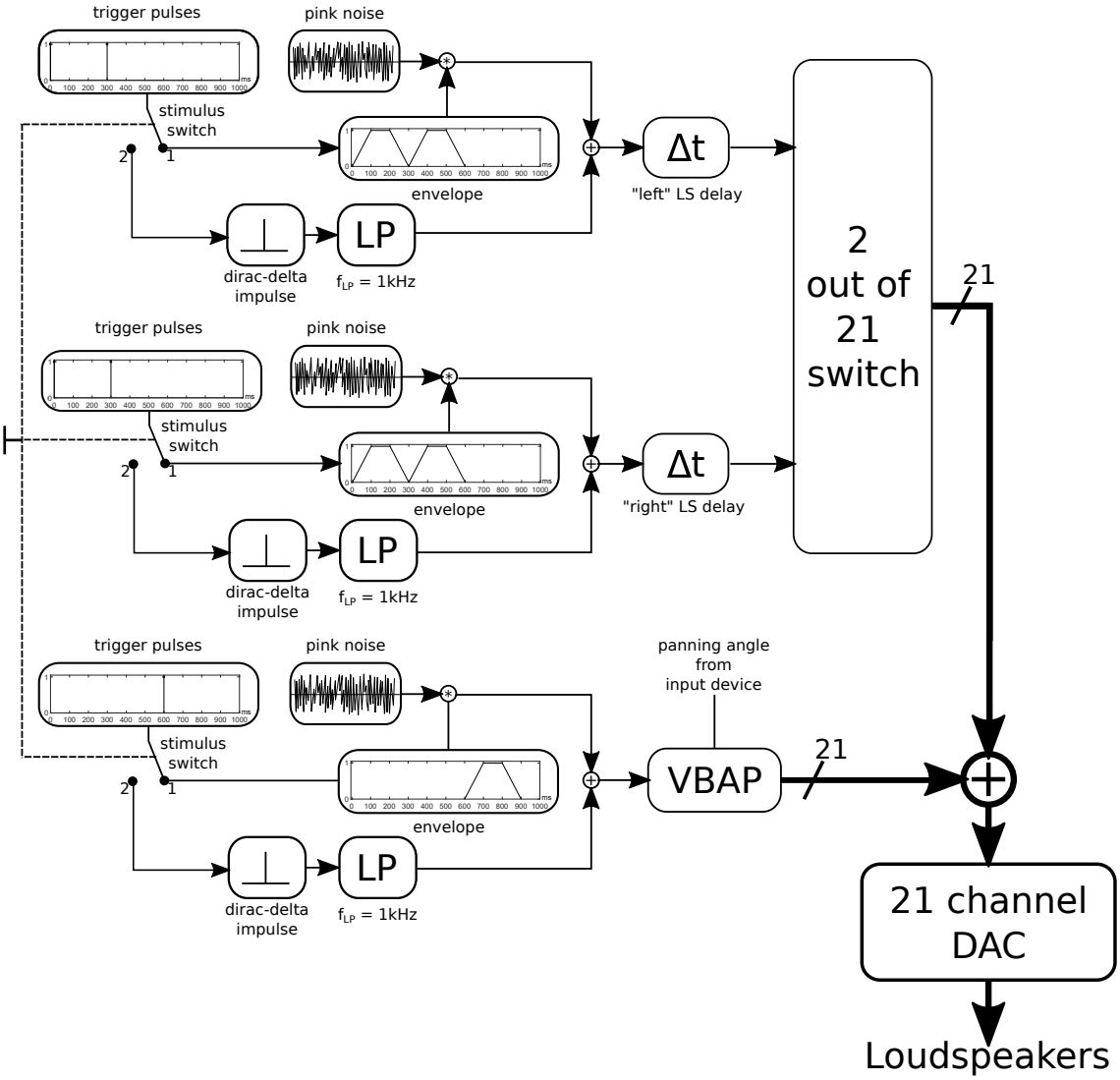


Figure 31 – Block diagram for generation of the stimuli.

The subjects had to use a combined shuttle/jog wheel with five additional buttons (see Figure 32) as input device. This device was also connected to the PC via USB. All buttons and wheels of the input device were assigned to specific functions during the listening experiment. By turning the outer shuttle wheel to the left/right against a rotational resistance, the third burst/click began to pan counter-/clockwise on the LS circle. A larger displacement angle resulted in a faster panning movement. By letting the shuttle wheel loose the rotational resistance turned the wheel back into the initial position and the third burst/click stopped to pan. With the inner jog wheel, the panning angle could be fine-tuned. Rotation counter-/clockwise resulted in a corresponding panning movement. The most left button, which is labeled with „Start“, was pressed to start the listening experiment. By pressing the „Answer“-button the adjusted panning angle was confirmed. The „Echo“-button also confirmed the panning angle but an additional flag for an echo detection respectively a source splitting was set. With the „next“-button, playback of the next stimulus triplet could be started after a valid confirmation. The most right button, which was labelled with „0“, could be used to reset the position of the third stimulus of the triplet to its initial position.

Figure 32 – Input device for the listening experiment.

4.3 Method

At first the supervisor loaded a test stimulus triplet series into the listening experiment Pd patch and explained the listening experiment to the subject with the help of a short trial run. The subjects were facing the direction of LS (1) and slight head movements during the listening test were allowed. The occurring stimulus triplets were presented and the input device was explained to the subject. After this short trial run the examiner loaded the personalized stimulus triplet series into the Pd patch, primed the listening test and leaved the experimental studio. The stimulus triplet series was randomized to compensate for signs of fatigue and order effects. The whole listening test could be started and controlled by the subject her-/himself. After pressing the „Start"-button on the input device a short speech recording „Listening test started!" indicated the start of the listening test. For playback of the first stimulus triplet loop the „next"-button had to be pressed. The stimulus triplet loop consisted of the two fixed position stimuli followed by the movable stimulus. The loop was repeated until the subject pressed the „Answer"-button. The random panning position of the two fixed position stimuli was dependent on the selected LS pair and the ICTD Δt between the two LSs. Playback of the movable stimulus always started at the center position (LS (1)). The subject listened to the stimulus triplet and had to pan the third stimulus so that it matched with the panning position of the first two stimuli. For the case that the subject was confused by the playback and panning situation he/she always had the possibility to pan the third stimulus back to the center position via the „0"-button. With the „Answer"-button the subject could confirm the adjusted panning angle of the third stimulus. During the listening test, there was the possibility to perceive an echo caused by large ICTDs, especially for $\pm$20ms and $\pm$30ms (cf. fig. 28). For this case the subject had to pan the third stimulus of the triplet to the position of the leading part of the first two stimuli and use the „Echo"-button to confirm its adjustment. After a confirmation via the „Answer"- or „Echo"-button the „next"-button had to be pressed to start playback of the following stimuli triplet loop. When „next" was pressed after the last confirmation, music was played to indicate the end of the listening test. Also the generated test data was saved to the PC and the supervisor was informed about the end of the listening test. Figure 33 shows a subject sitting inside the setup during the listening experiment.

Figure 33 – Setup of the listening experiment with subject inside.

Each subject evaluated the 168 stimuli triplets and spent about 48 up to 83 minutes. Average test duration was approximately 69 minutes. The experiment took place from 15th to 18th of March 2018 at the experimental studio CUBE of the Institute of Electronic Music and Acoustics in Graz, Austria. 16 subjects (4 females and 14 males) with an average age of 26 years participated in the listening test.

4.4 Results

The evaluation of the panning angles φ and echo detections, which where indicated by the subjects, focusses on three topics:

- localization depending on ICTD,
- determination of an optimal slope parameter τ,
- finding of an echo detection resp. source splitting threshold Δt_{echo}.

All the processing of the listening test data is done in MATLAB. At first the listening experiment data of every subject from the generated text files is read into a single matrix. Afterwards the randomized data is sorted and split up into frontal and lateral playback directions for further processing.

• Frontal listening test data:

To check if the subjects made reliable panning angle indications, the cases with single LS playback are used. For the center LS (1) a threshold of $\Delta\varphi = \pm 5°$ is defined for exclusion of the data set of a subject. Every subject could satisfy this requirement for playback of PPN as well as clicks. Figure 34 shows the median value and the corresponding interquartile range (IQR) for all panning angles φ for playback of pink pulsed noise

(a) and clicks (b) per subject and for all subjects. The dashed line indicates the center playback direction. For playback of PPN (fig. 34 (a)) we can see that some subjects tend to localize the virtual sound source more on the right side. However, the summarized data of all subjects has a $\overline{\varphi}_{all} \approx 0°$ but the $IQR_{all} \approx 25°$ tends also to the right side. For playback of clicks (fig. 34 (b)) the $IQRs$ are larger ($IQR_{all} \approx 45°$) but some subjects (3,4, and 13) localized the virtual sound source for all $ICTDs$ at the center position. Again for all data $\overline{\varphi}_{all} \approx 0°$ and IQR_{all} tends to the right. Based on this analysis all the data sets could be used for evaluation. In the next processing step indicated panning angles $\pm\varphi$ that are bigger than the aperture $\pm\alpha$ of the respective LS pair that was used for playback are set to $\pm\alpha$. Especially this is needed for a large aperture α because of worse localization at lateral playback directions. This processing step is equivalent to a restriction of $\pm\varphi \lessgtr \pm\alpha$ in the listening test. Moreover, the data is checked for the right panning angle when an echo is detected. If φ is not panned in direction of the leading LS for an echo case, the sign of φ is inverted.

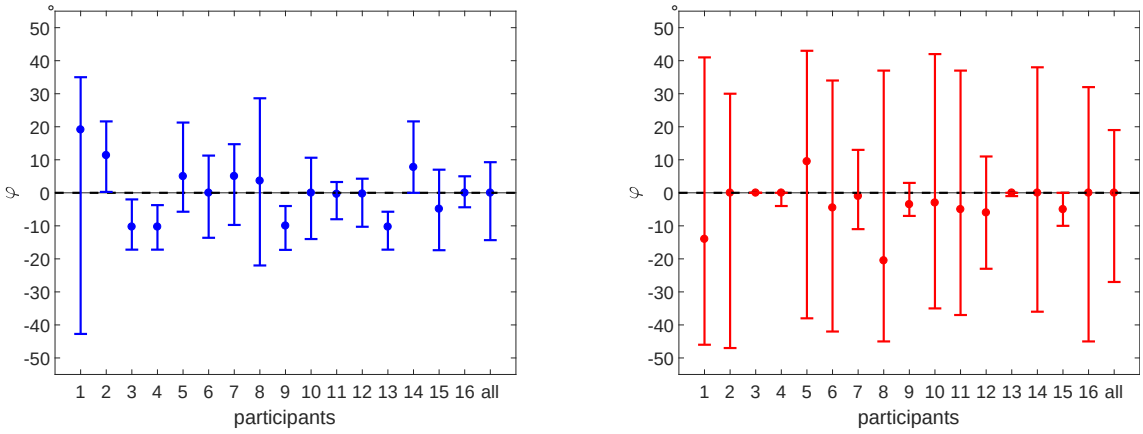


Figure 34 – Median values and IQR for all φ per subject and for all subjects. Frontal LS pairs with PPN (a) and clicks (b).

• Lateral listening test data:

The data of playback from left and right LS pairs is combined for every subject. This is done according to tab. 3 for even and odd numbered subjects. Finally, a data set with playback from lateral LS pairs (covering all apertures α) per subject is provided. Again an analysis of median value $\overline{\varphi}$ and corresponding IQR is done per subject and for all subjects (fig. 35). For playback of PPN in general the IQR is larger in comparison to the frontal playback direction ($IQR_{all} \approx 40°$). Furthermore, most subjects tended to localize the virtual sound source more from the front ($\overline{\varphi}_{all} \approx 85°$). It seems that the front/back-confusion has critical influence on the localization for lateral playback directions. This effect is even more distinct for playback of clicks ($\overline{\varphi}_{all} \approx 78°$). The IQR for lateral playback of clicks is similar to this for frontal playback. Despite the mentioned front/back-confusion, all the data is used for analysis. Further processing is done similar to that from the frontal listening test data.

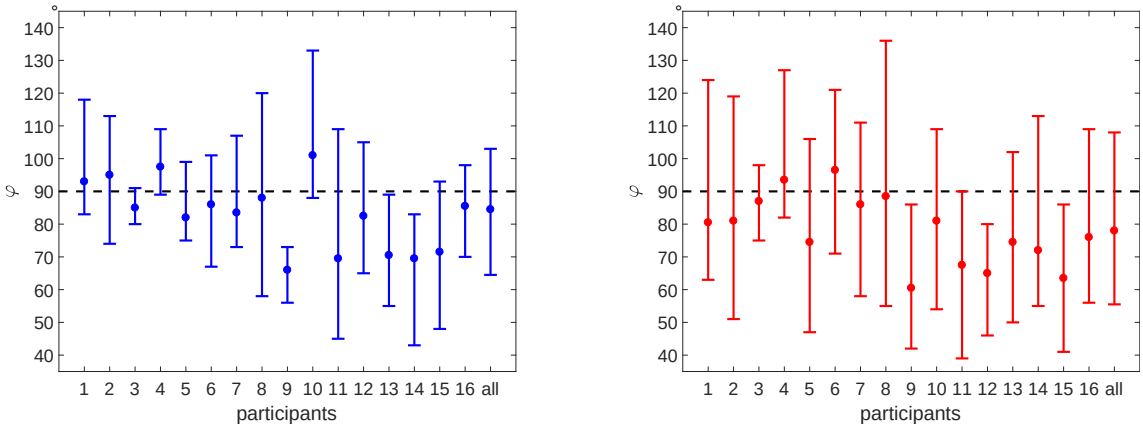


Figure 35 – Median values and IQR for all φ per subject and for all subjects. Lateral LS pairs with PPN (a) and clicks (b).

4.4.1 Localization depending on ICTD

Figures 36 and 37 are showing the localization direction depending on the ICTD Δt for every frontal LS pair in a respective polar plot. Plots from fig. 36 are showing the results for playback of PPN. Accordingly, fig. 37 shows results for click playback. The active LS pair is coloured black in every plot. For every ICTD Δt, the median panning angle $\overline{\varphi}$ and the corresponding 95 % confidence interval is plotted. Note that because of the polar plot, same size confidence intervals appear larger for a larger ICTD. For reasons of better overview only positive ICTDs are plotted. Therefore, a positive $\overline{\varphi}$ for a certain ICTD indicates that this ICTD is produced by a delaying the right LS and vice versa. Linear interpolation is performed between all $\overline{\varphi}(\Delta t)$. Moreover, a green dot for $\overline{\varphi}$ indicates that there is an echo detection very likely according to sec. 4.4.2.

For playback of PPN a larger ICTD results in a larger displacement of the indicated playback direction for the same LS aperture angle α. However even for a large Δt the panning angle $\overline{\varphi}$ never reaches the LS direction of a playing pair. Furthermore, a higher uncertainty at indication because of larger confidence intervals can be observed for a larger Δt. Confidence intervals for the same delay at the left and right LS overlap quite often, especially for small ICTDs. This means that the direction indications for the virtual sound source do not differ significantly from each other and direction indication is not consistent. In fig. 36 (d) the source displacement for a delayed left LS is significantly smaller then for a delayed right LS. Moreover for a $\Delta t = 5$ ms $\overline{\varphi}$ is positive. Maybe this quite unexpected behaviour is caused by colouration effects and/or unwanted reflections produced by the positioning of the LS in the room. The echo detection is symmetric for a delay at the left and right LS of a pair and mainly happens for a large $\Delta t = 20$ ms resp. 30 ms.

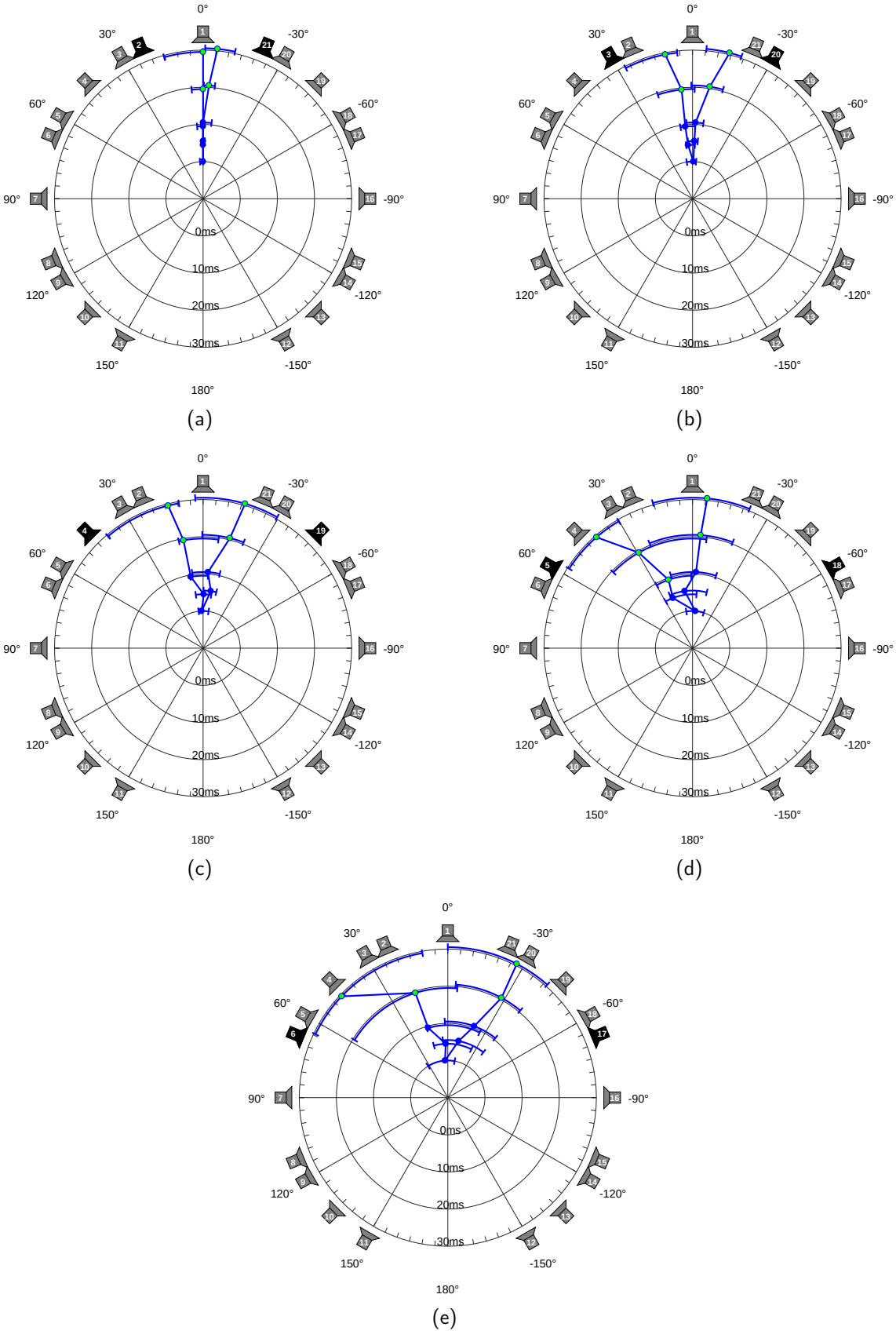


Figure 36 – Median values and 95% confidence intervals for PPN. Frontal LS pairs (a): (2,21), $\alpha = 45°$, (b): (3,20), $\alpha = 60°$, (c): (4,19), $\alpha = 90°$, (d): (5,18), $\alpha = 120°$, (e): (6,17), $\alpha = 135°$.

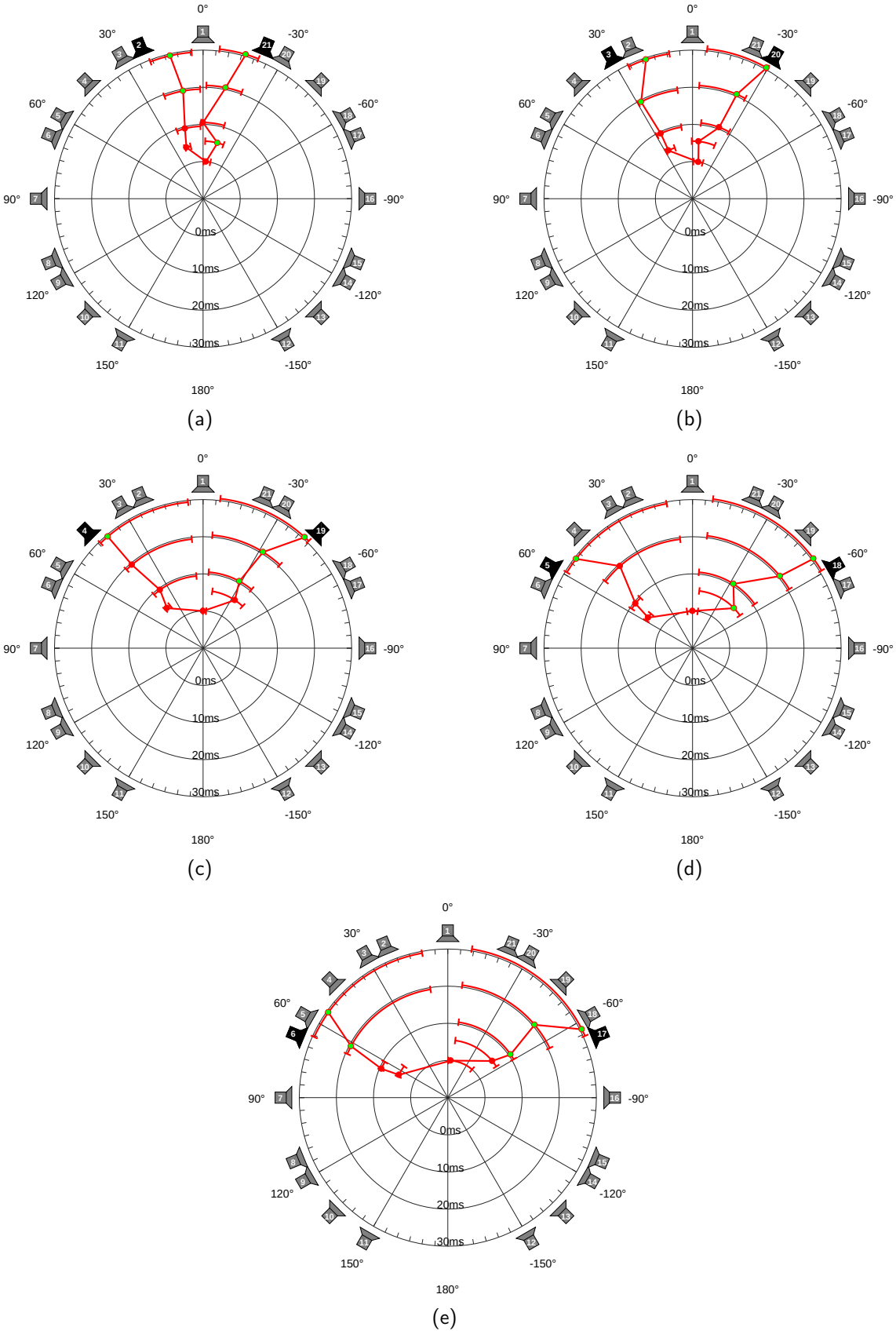
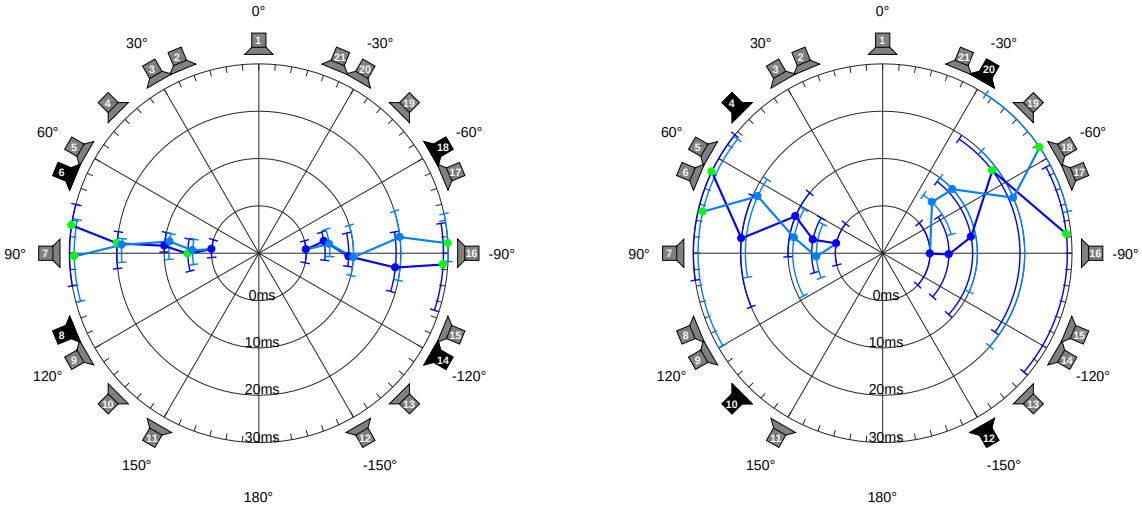
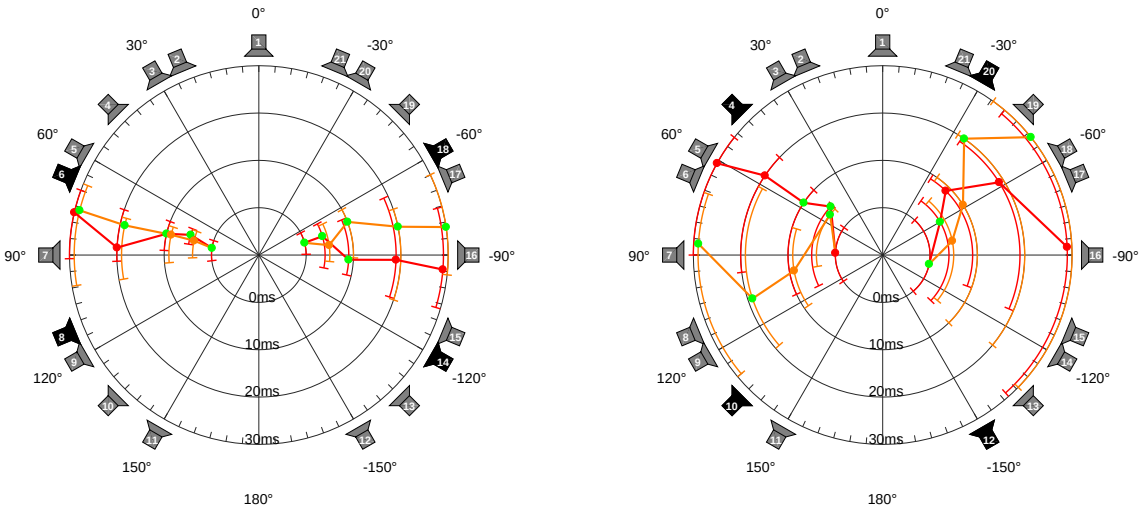


Figure 37 – Median values and 95% confidence intervals for clicks. Frontal LS pairs (a): (2,21), $\alpha = 45°$ (b): (3,20), $\alpha = 60°$ (c): (4,19), $\alpha = 90°$, (d): (5,18), $\alpha = 120°$, (e): (6,17), $\alpha = 135°$.



(a) (b)

Figure 38 – Median values and 95% confidence intervals for PPN. Lateral LS pairs (a): (8,6)/(18,14), $\alpha = 45°$ and (b): (10,4)/(20,12), $\alpha = 60°$.



(a) (b)

Figure 39 – Median values and 95% confidence intervals for clicks. Lateral LS pairs (a): (8,6)/(18,14), $\alpha = 90°$ and (b): (10,4)/(20,12), $\alpha = 120°$.

For playback of clicks, also larger ICTDs lead to larger displacements and larger confidence intervals but even more distinct. Although the variation of direction indication is higher, confidence intervals do not overlap. Consequently there are significant different direction indications for the same Δt at left/right LS and therefore localization is more robust. Moreover for large aperture angles α (LS pairs (4,19), (5,18) and (6,17)) direction indications nearly reach the LS directions. Conspicuous behaviour can be observed for echo detection. Subjects can detect more often echoes for a delay at the left LS

(e.g. at fig. 37 (c) and (d)). Also for a small $\Delta t = 5$ ms echo detection is likely. This result is expected because of a smaller echo threshold for more transient signals.

Figures 38 and 39 are showing $\overline{\varphi}(\Delta t)$ for the lateral LS pairs. The depiction of the results is done in a similar way to fig. 36 and 37. In every plot there are results for two LS pairs depicted. Moreover each side of a panning curve is coloured differently. A blue/red colouration indicates a delay at the counter-clockwise displaced LS (in comparison to the center position of the LS pair), a light blue/orange colouration a delay at the clockwise displaced LS. This is done because of a strong overlapping of the φ data. This behaviour is probably caused by the front/back-confusion resp. effects from the cone-of-confusion and occurs despite small head movements were allowed. For playback of PPN over LS pairs (8,6), (18,14) and (10,4) results for positive and negative ICTD are quite identical. Only for the large aperture angle $\alpha = 120°$ of LS pair (20,12) it could be possible to detect a tendency of localization in back direction for a delay at LS (20). Although the panning angle indication for $\Delta t = 20$ ms and large confidence intervals contradict this impression. In general, confidence intervals are increasing with an ascending Δt and α. Similar to the evaluation of the frontal LS pairs, localization is slightly better for playback of clicks. For LS pair (8,6) again positive and negative ICTD lead to similar panning angles. For the LS pairs (18,14) and (10,4) there can be two main panning directions assumed. Nevertheless, they are not significantly different from each other. The large aperture angle of LS pair (20,12) is leads to very large confidence intervals. Localization is very inconsistent for this case.

Summarizing it could be shown that localization based on ICTD is quite possible for frontal playback directions. Lateralization becomes more distinct for more transient signals. For increasing aperture angles α and an increasing Δt, the accuracy of the localization decreases. For lateral playback directions localization based only on ICTD is quite questionable. In general, subjects tend to localize the virtual sound source more from the front. Also localization is very inconsistent, especially for a large aperture angle α of the used LS pair. This behaviour is mainly caused by the front/back-confusion.

4.4.2 Echo detection

As mentioned in subsection 4.3, it was possible that the subjects recognized an echo in consequence of large ICTDs. In this case, the subjects were asked to confirm their answers with the „Echo“-button and a flag in the listening test data was set. Figures 40 and 41 are showing the echo detection rate (EDR) depending on the ICTD Δt for the frontal LS pairs. It is plotted for all subjects and for playback of PPN and clicks. Also the combined EDR for all center LS pairs is shown on position (f) in each figure. Combination is possible because an analysis of variance (ANOVA) performed with the MATLAB function `anova1` is leading to $p_{noise} = 0.69$ and $p_{clicks} = 0.90$. Moreover the mean echo detection rate ($\overline{EDR}$) is plotted (red dashed line). When comparing the EDR for PPN and clicks, it is noticeable that the clicks cause a higher EDR. This observation confirms the fact that more transient resp. impulsive signals tend to a lower echo threshold (cf. subsection 3.3). The $\overline{EDR}$ with 38.61 $\%$ for clicks is significant higher than for PPN with 25.69 $\%$.

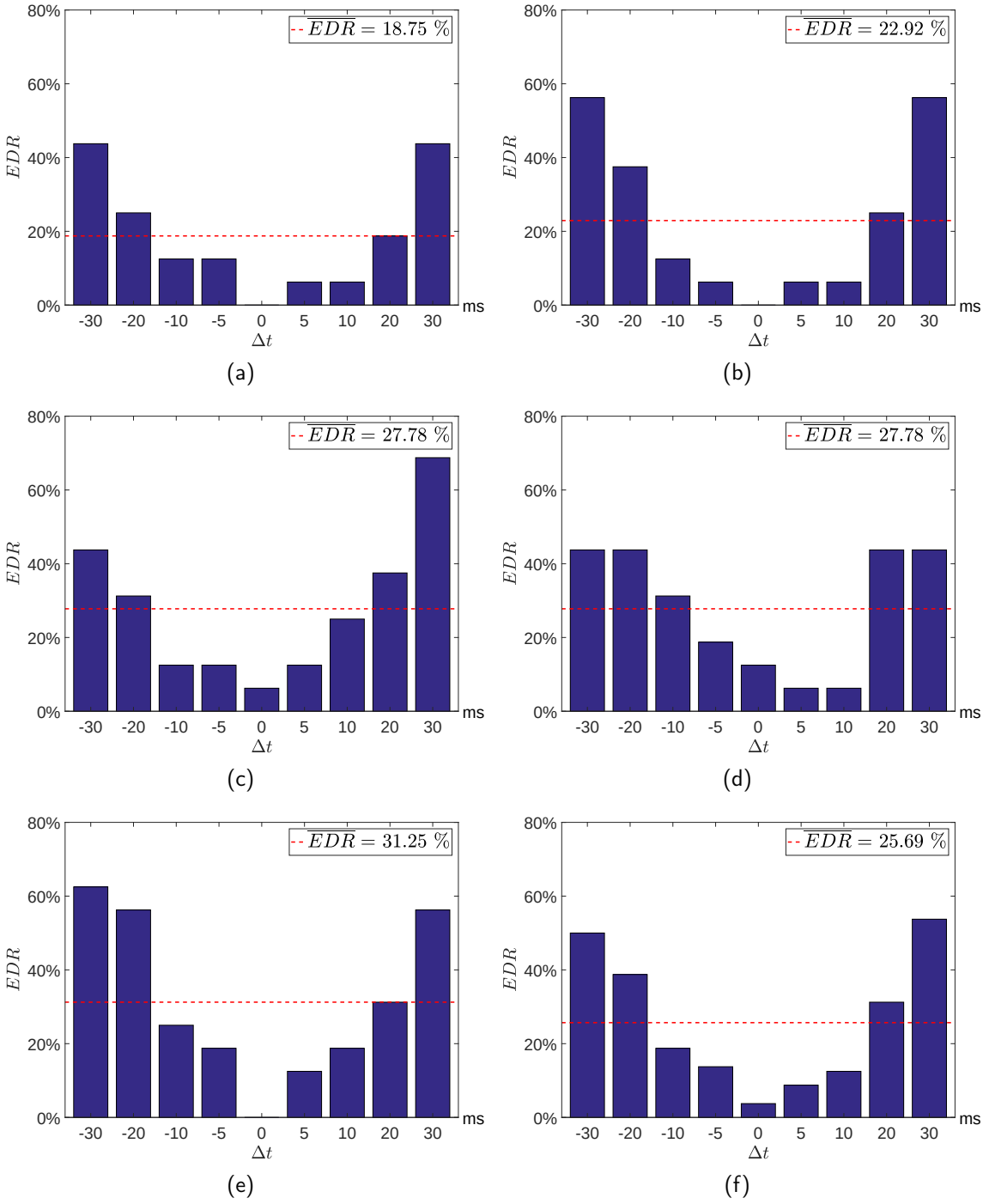


Figure 40 – EDR for occuring ICTD Δt with PPN. Frontal LS pairs (a): (2,21), $\alpha = 45°$, (b): (3,20), $\alpha = 60°$, (c): (4,19), $\alpha = 90°$, (d): (5,18), $\alpha = 120°$, (e): (6,17), $\alpha = 135°$ and (f): all pairs.

Also the statement from subsection 3.3 that direction of arrival of the lagging signal influences the echo detection can be confirmed. For the more narrow LS pairs the $\overline{EDR}$ ((2,21): 18.75 % for PPN and 32.64 % for clicks) is smaller then for the wider ones

((6,17): 31.25 % and 42.36 %). Symmetry for positive and negative delays is given for PPN playback. Interestingly, symmetry is weaker for the clicks because there are more echo detections for the positive delay. Maybe some room reflections caused this perception.

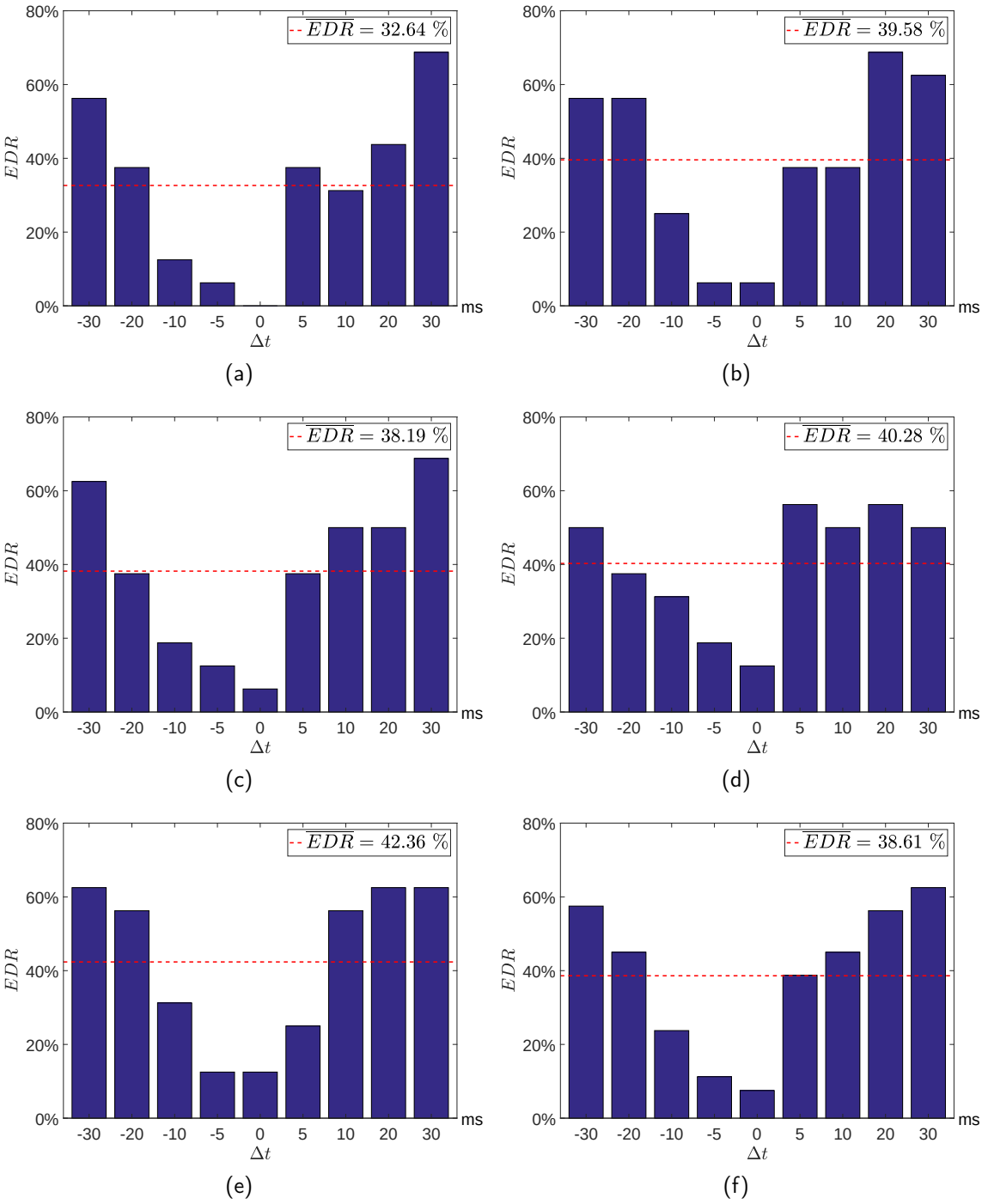


Figure 41 – EDR for occurring time delay values with clicks. Frontal LS pairs (a): (2,21), $\alpha = 45°$, (b): (3,20), $\alpha = 60°$, (c): (4,19), $\alpha = 90°$, (d): (5,18), $\alpha = 120°$, (e): (6,17), $\alpha = 135°$ and (f): all pairs.

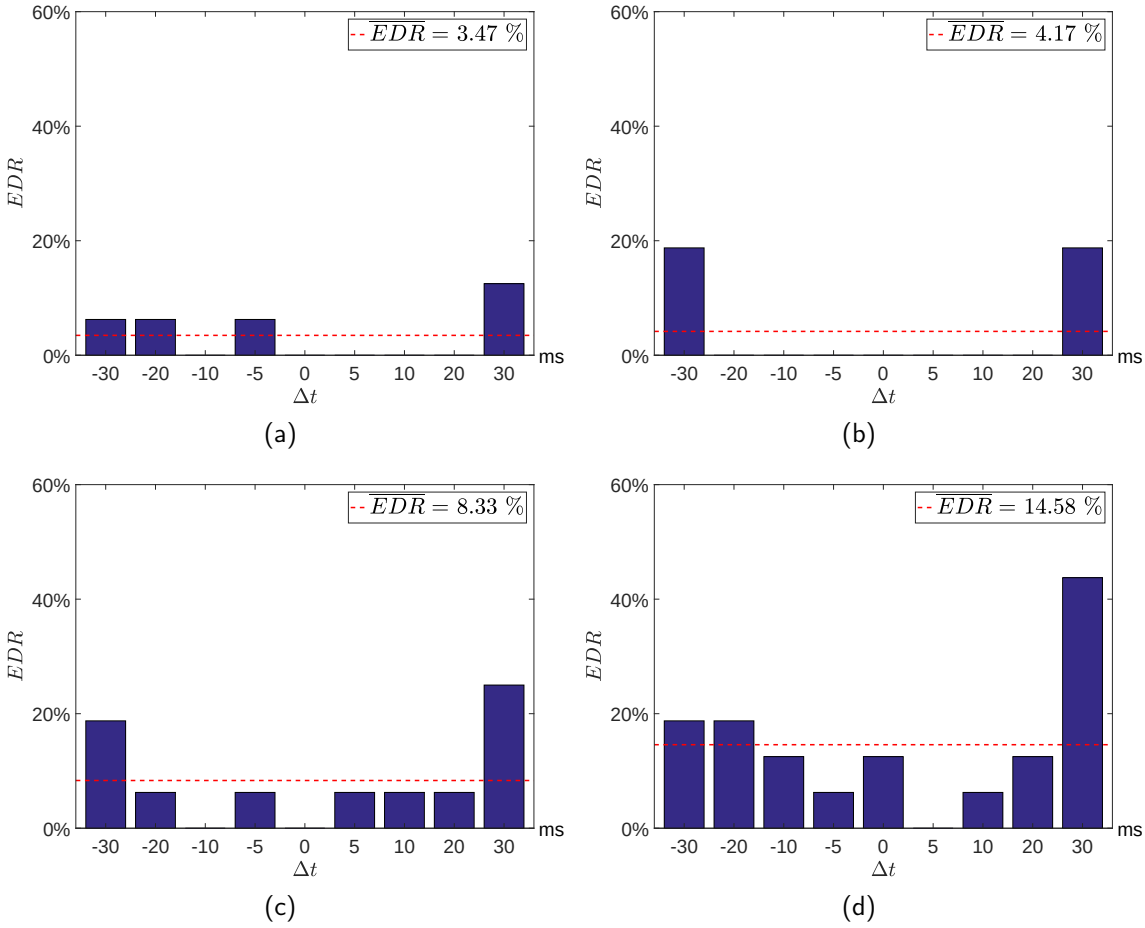


Figure 42 – EDR for occurring ICTD Δt with PPN. Lateral LS pairs (a): (8,6)/(17,15), $\alpha = 45°$, (b): (9,5)/(18,14), $\alpha = 60°$, (c): (10,4)/(19,13), $\alpha = 90°$ and (d): (11,3)/(20,12), $\alpha = 120°$.

Also fig. 42 and 43 show EDR and $\overline{EDR}$ (red dashed line) of all subjects for the lateral LS pairs. The data for left and right playback is summed up for LS pairs with the same aperture angle. The data for the right-hand LS pairs was processed so that a positive sign indicates a delay for the LS that was in the frontal area (as for the left-hand pairs). Combination of the EDR distributions for PPN playback was not possible because of significant differences between the data ($p_{noise} = 0.0455$). For click playback it is $p_{clicks} = 0.57$. In comparison to the center LS pairs EDR is significantly lower. $\overline{EDR}$ for all pairs for clicks is 20.31 %. Also the fact that a more transient signal causes a lower echo threshold is confirmed again. More narrow LS pairs also have a lower EDR (cf. fig. 42 (a) with (d) and fig. 43 (a) with (d)). For the playback of clicks at lateral positions the EDR distribution is not symmetric anymore. In general, the subjects can perceive an echo in the front better than in the back. It seems that the symmetry of the EDR histograms is shifted to positive delays. In consequence many subjects perceived an echo also for an ICTD of 0 ms (cf. fig. 43 (a) and (d)).

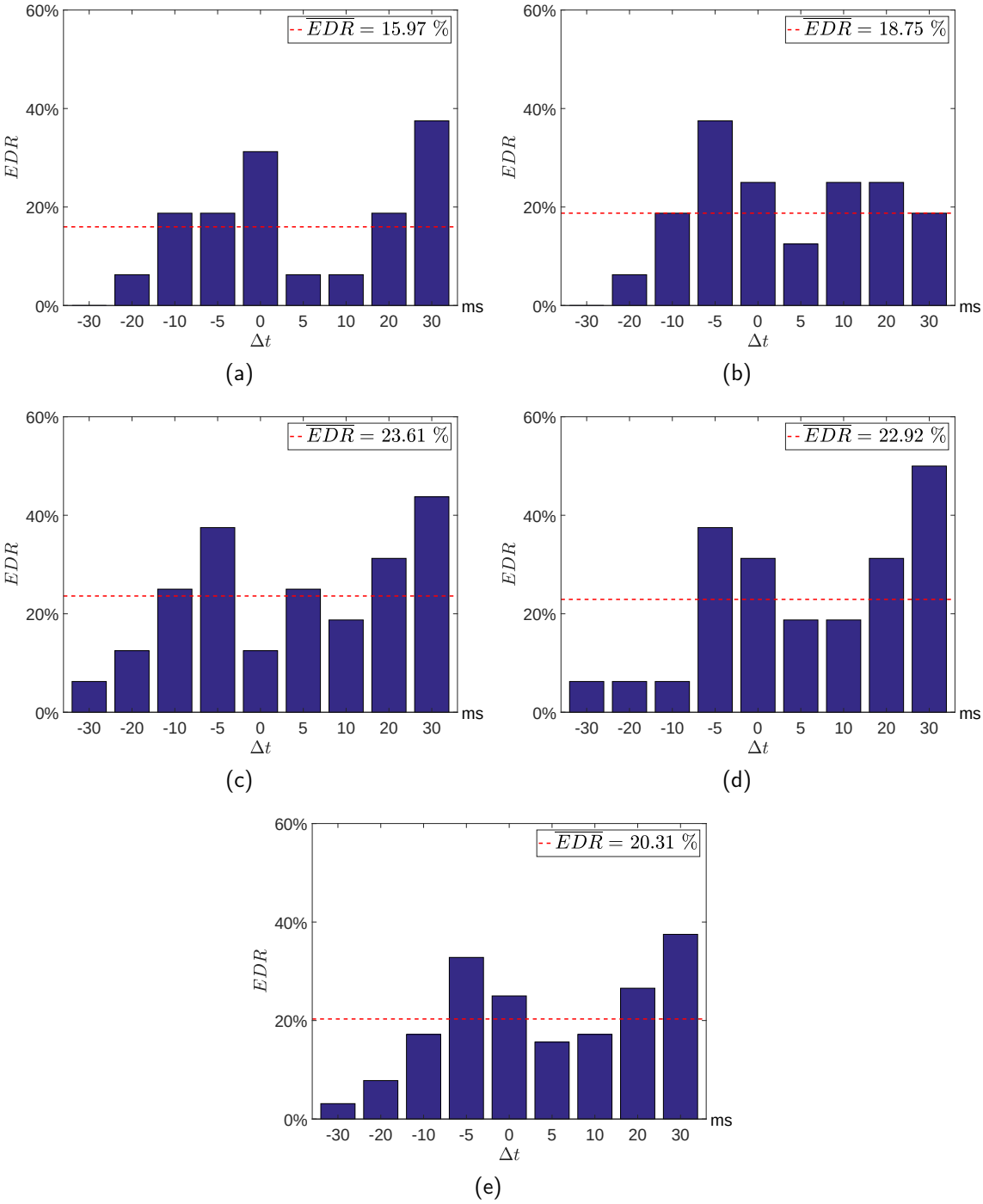


Figure 43 – EDR for occurring ICTD Δt with clicks. Lateral LS pairs (a): (8,6)/(17,15), $\alpha = 45°$, (b): (9,5)/(18,14), $\alpha = 60°$, (c): (10,4)/(19,13), $\alpha = 90°$, (d): (11,3)/(20,12), $\alpha = 120°$ and (e): all pairs.

When it comes to reproduction of a three-dimensional soundfield with a spherical LS array the occurring of an echo can also be seen as the result of source splitting. Source splitting is the effect when a virtual sound source that is produced by several LSs splits into 2 or more auditory objects because of unfavourable ICLD and/or ICTD. In order to get a certain threshold for the detection of an echo resp. source splitting, it is required to define a measure that creates a relation between the probability of the echo detection rate (EDR) and the ICTD Δt. To get a more precise curve for the relation of EDR and Δt, it is advantageous to combine the data for positive and negative ICTDs. For combination, we first have to check if the EDR distributions for positive and negative delays are not significantly different. This can be done by a t-test for the two sides of every EDR distribution. Tables 5 and 6 are showing the p-values for this comparison for the frontal and lateral LS pairs. When playing back clicks at the center LS pair (5,18) ($p_{clicks} = 0.04$) left and right side of the EDR distribution are significantly different from each other. Neglecting this fact, we combine the data for positive and negative delays for all LS pairs.

LSP-Pair	(2,21)	(3,20)	(4,19)	(5,18)	(6,17)	all
α	45°	60°	90°	120°	135°	-
p_{noise}	0.53	0.65	0.40	0.28	0.37	0.64
p_{clicks}	0.18	0.25	0.11	0.04	0.36	0.15

Table 5 – p-values for the comparison of the EDR distributions for positive and negative delays at the frontal LS pairs.

LSP-Pair	(6,8)/(15,17)	(5,9)/(14,18)	(4,10)/(13,19)	(3,11)/(12,20)	all
α	45°	60°	90°	120°	-
p_{noise}	0.26	1.00	0.52	0.77	0.90
p_{clicks}	0.81	0.67	0.43	0.16	0.36

Table 6 – p-values for the comparison of the EDR distributions for positive and negative delays at the lateral LS pairs.

Also combination of the EDR distributions of all LS pairs for frontal ($p_{noise} = 0.90$, $p_{clicks} = 0.96$) and lateral ($p_{noise} = 0.15$, $p_{clicks} = 0.76$) playback is possible. A curve fitting can be applied. For this purpose an exponential function model for PPN and a logarithmic one for clicks are used. Both functions seem to be suitable for the resp. data. Moreover both curves are strict monotonically increasing or with Δt. Curve fitting is done in the least-squares sense using the MATLAB function `lsqcurvefit`. This function returns the coefficients $p(\Delta t)$ that are a best fit for the given EDR at a certain Δt. Both can be written in the following way:

$$\widehat{EDR}(\Delta t)_{PPN} = p_1 \cdot \exp(p_2 \Delta t \frac{1}{\text{ms}} + p_3), \tag{38}$$

$$\widehat{EDR}(\Delta t)_{clicks} = p_1 \cdot \ln(p_2 \Delta t \frac{1}{\text{ms}} + p_3). \tag{39}$$

For determining a minimum threshold value Δt_{echo}, at which echo perception is most likely, the intersection of $p(\Delta t)$ and an $EDR = 50\%$ is chosen and projected on the ICTD Δt.

Figures 44 and 45 are showing the half-sided EDR distributions for both signal types for the center LS pairs and their combination (f). It is evident that the mean echo detection rate ($\overline{EDR}$) is increasing with the aperture angle α of the LS pair. Moreover, Δt for intersection of $\overline{EDR}$ with the polynomial curve fit $\widehat{EDR}(\Delta t)_{poly}$ is decreasing for an increasing α. This results strengthens the statement (III) from subsection 3.3. Furthermore, playback of clicks is leading to a lower Δt at intersection and higher $\overline{EDR}$ then PPN (cf. fig. 44 (f) with fig. 45 (f)). This strengthens statement (I) from subsection 3.3. The values for Δt_{echo} also underline this fact. EDR distributions for the lateral LS pairs are more ambiguous. For PPN (fig. 46) a rising α also leads to a higher $\overline{EDR}$ but it barely has influence on Δt at intersection. Moreover Δt at intersection is higher then for the frontal LS pairs (cf. fig. 44 (f) with fig. 46 (e)). This also holds for playback of clicks (cf. fig. 47). Also for lateral playback, clicks lead to a lower Δt at intersection. Because of the generally low EDR for lateral playback, a computation of Δt_{echo} for the combined data of front and back is not possible (cf. fig. 46 and 47 (e)). Unfortunately, it is not possible to combine the EDR distributions for frontal and lateral playback. The ANOVA for PPN (with EDR distributions from fig. 44 (f) and fig. 46 (e)) is resulting in a $p_{noise} = 0.0231$ and for clicks (with EDR distributions from fig. 45 (f) and fig. 47 (e)) in a $p_{clicks} = 0.0462$.

Table 7 holds the depending parameters which results in the curves depicted in fig. 48.

direction	signal	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{p_3}$
frontal	PPN	0.0918	0.0605	-0.0451
	clicks	0.325	0.1701	1.2639
lateral	PPN	0.0124	0.0927	-0.0033
	clicks	0.1942	0.1417	1.1576

Table 7 – Parameters p_1, p_2 and p_3 for eq. 38 and 39 describing the curves from fig. 48.

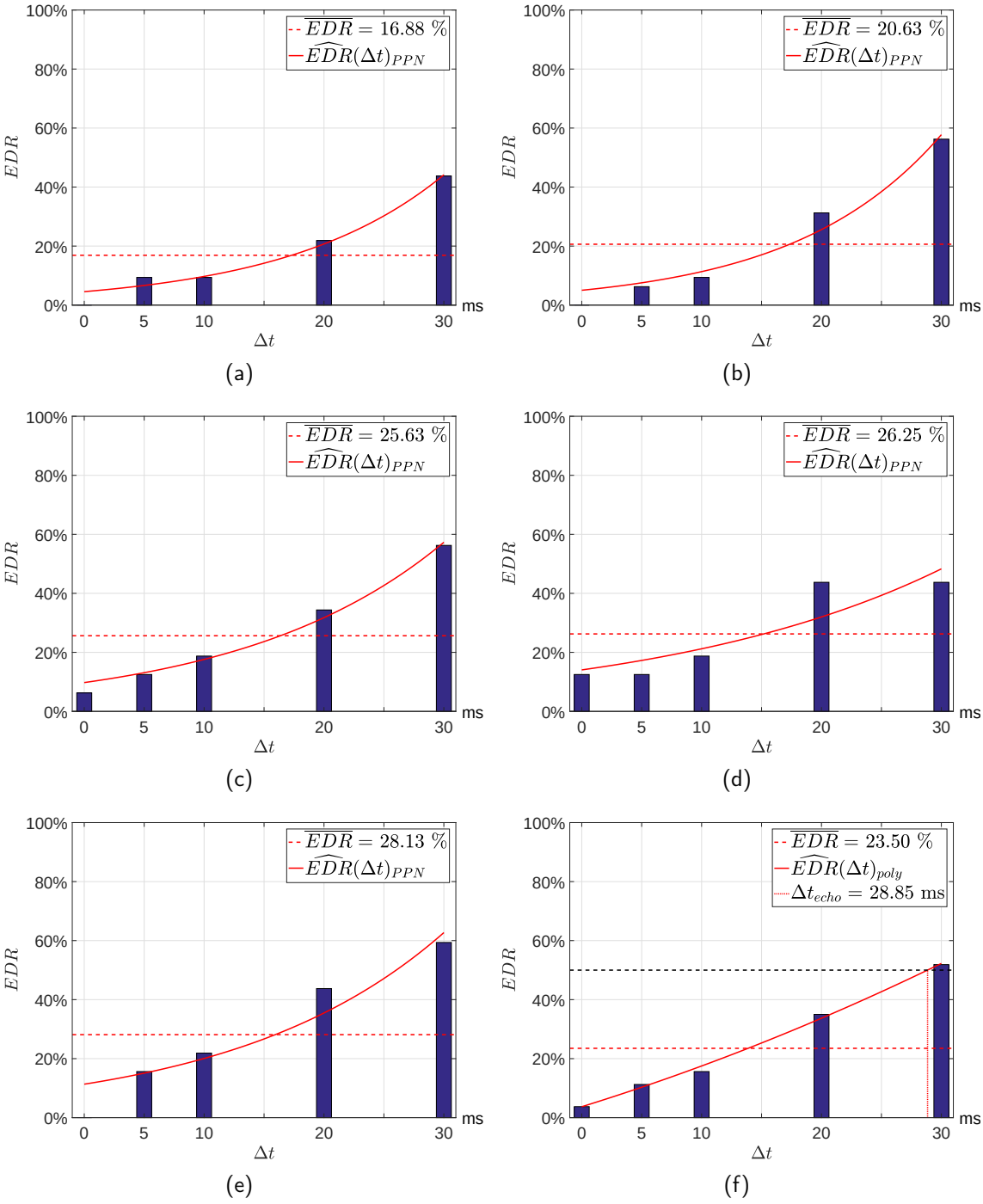


Figure 44 – EDR for occurring ICTD Δt with PPN for frontal LS pairs ((a): (2,21), $\alpha = 45°$, (b): (3,20), $\alpha = 60°$ (c): (4,19), $\alpha = 90°$ (d): (5,18), $\alpha = 120°$ (e): (6,17), $\alpha = 135°$ and (f): all pairs) with mean $\overline{EDR}$ and a polynomial curve $\widehat{EDR}(\Delta t)_{PPN}$.

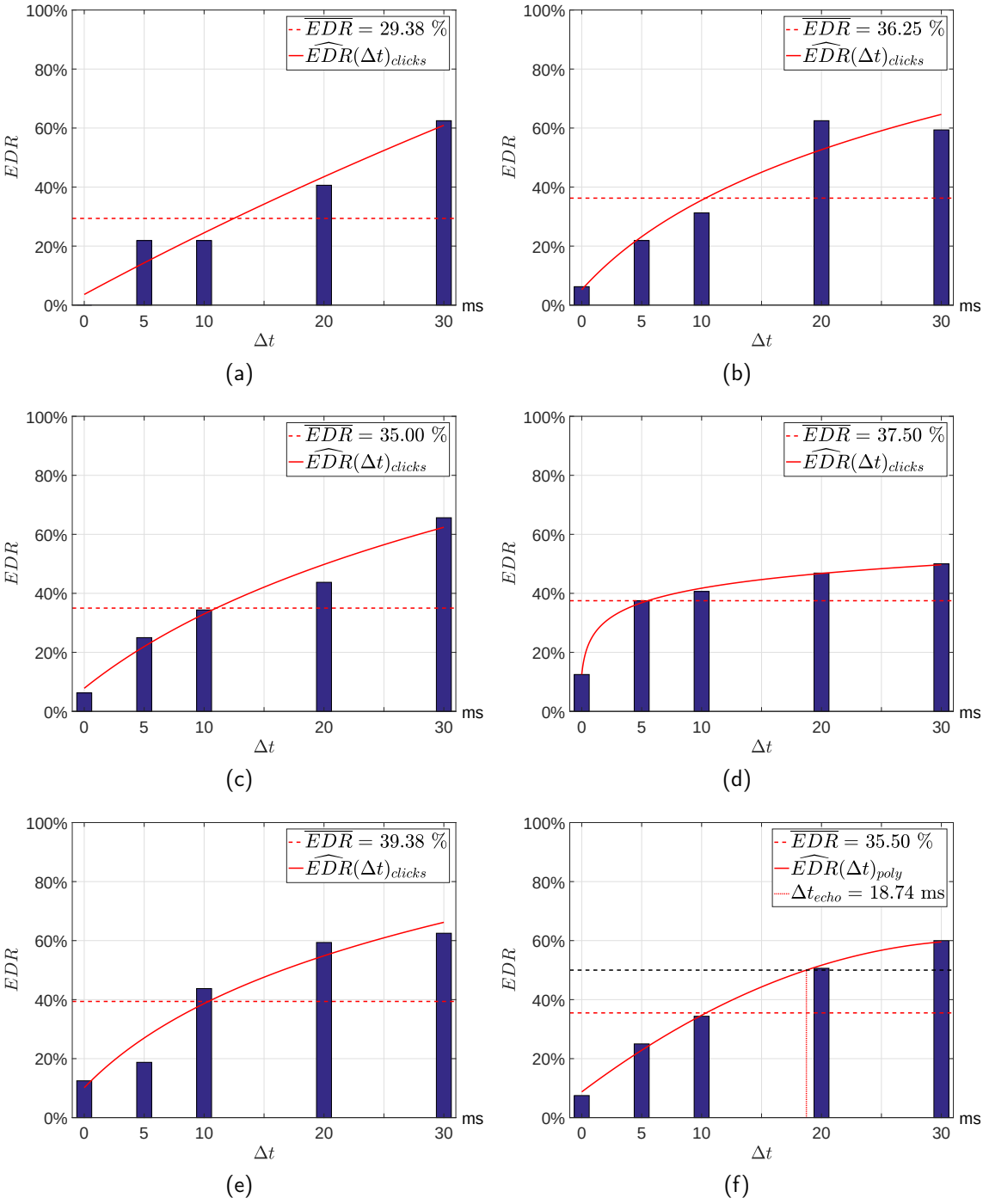


Figure 45 – EDR for occurring ICTD Δt with clicks for frontal LS pairs ((a): (2,21), $\alpha = 45°$, (b): (3,20), $\alpha = 60°$ (c): (4,19), $\alpha = 90°$ (d): (5,18), $\alpha = 120°$ (e): (6,17), $\alpha = 135°$ and (f): all pairs) with mean $\overline{EDR}$ and a curve fit $\widehat{EDR}(\Delta t)_{clicks}$.

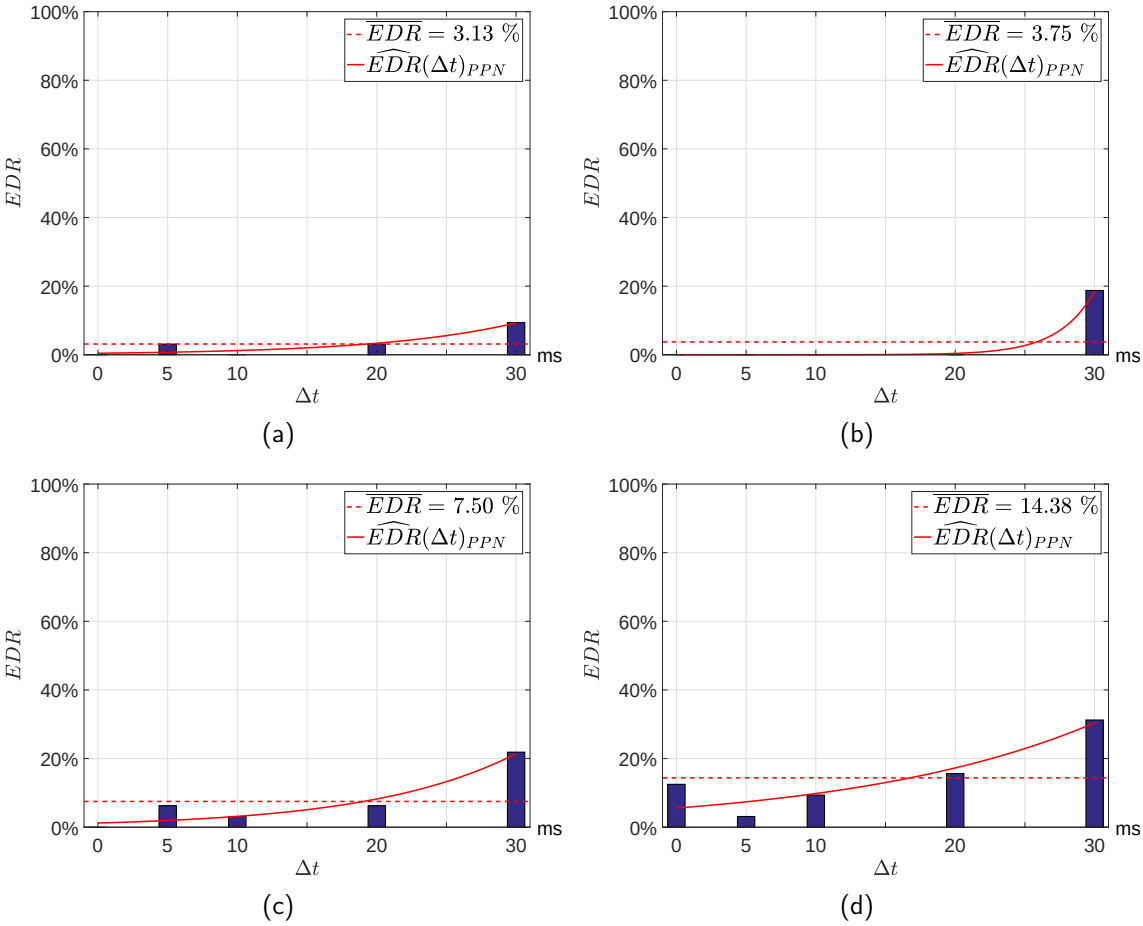


Figure 46 – EDR for occurring ICTD Δt with PPN for lateral LS pairs ((a): (8,6)/(17,15), $\alpha = 45°$, (b): (9,5)/(18,14), $\alpha = 60°$, (c): (10,4)/(19,13), $\alpha = 90°$ and (d): (11,3)/(20,12), $\alpha = 120°$) with mean $\overline{EDR}$ and a curve fit $\widehat{EDR}(\Delta t)_{PPN}$.

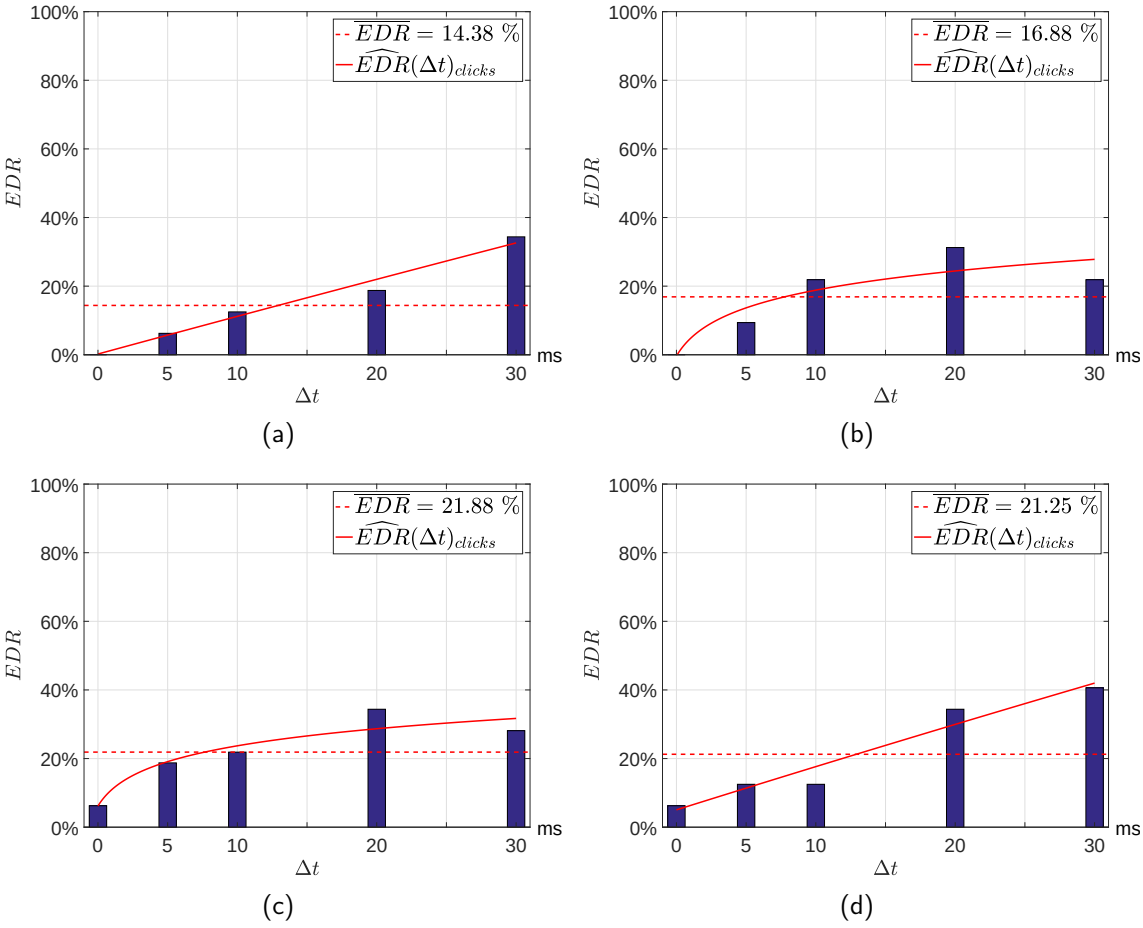


Figure 47 – EDR for occurring ICTD Δt with clicks for lateral LS pairs ((a): (8,6)/(17,15), $\alpha = 45°$, (b): (9,5)/(18,14), $\alpha = 60°$, (c): (10,4)/(19,13), $\alpha = 90°$ and (d): (11,3)/(20,12), $\alpha = 120°$) with mean $\overline{EDR}$ and a curve fit $\widehat{EDR}(\Delta t)_{clicks}$.

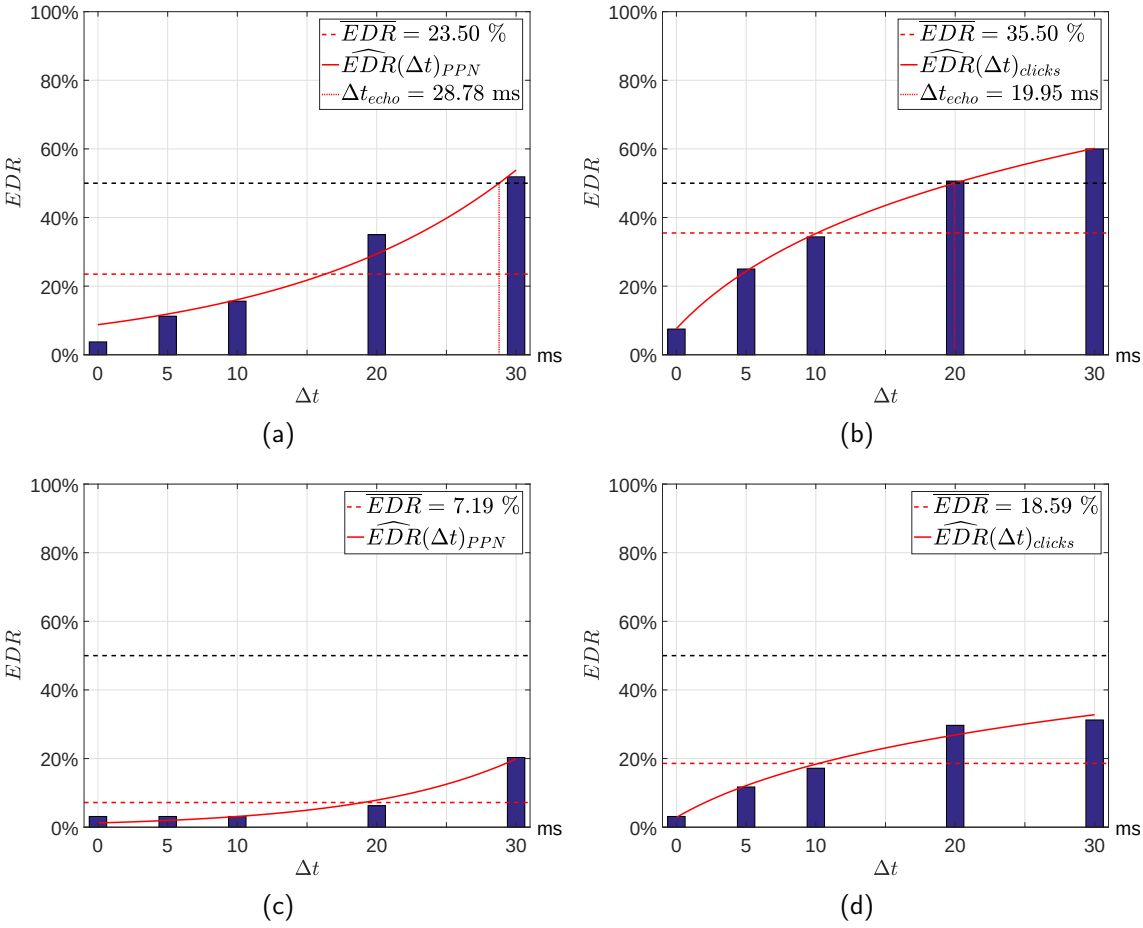


Figure 48 – EDR for occurring ICTD Δt. (a): frontal LS pairs with PPN, (b): frontal LS pairs with clicks, (c): lateral LS pairs with PPN and (d): lateral LS pairs with clicks.

Summarizing the results of this section the $\overline{EDR}$ increasing and consequently the Δt_{echo} decreases for more transient stimulus signals. Moreover, the directional difference between the leading and the lagging stimulus influences the $\overline{EDR}$. A larger aperture angle α leads to a higher $\overline{EDR}$ and a smaller Δt_{echo}. In general, echo discrimination in the frontal area is good (cf. fig. 48 (a) and (b)). For PPN a echo detection from $\Delta t_{echo,PPN} \geq 29$ ms is most likely. For clicks it is obviously smaller with $\Delta t_{echo,clicks} \geq 20$ ms. The two values are located within the intervals for an upper and lower echo threshold that was determined in [12] for horizontal and azimuthal playback situations with pink noise and speech (similarity to clicks possible). For lateral directions, echo discrimination becomes worse, especially for PPN ($\overline{EDR} < 10$ %). Furthermore for this stimulus signal, the detection is very inconsistent and a threshold for echo detection can not be determined. For lateral clicks the EDR is approx. twice as large but also Δt_{echo} can not be determined.

Because echo detection is accompanied with source-splitting in soundfield reproduction it is reasonable to integrate this knowledge into the $\mathbf{r_E}$-vector model. This can be done with the help of the detection threshold Δt_{echo}, which defines a time window after an

incoming acoustical signal that arrives at the listening position. If the delay between two identical signals is larger than Δt_{echo} source-splitting is most likely. Further developments regarding source-splitting can be found in section 5.3. In many common performance situations there is a main head orientation of the audience. Based on the results of this section, it is reasonable to define a frontal perceptive area where echo detection sensitivity is the highest. This sensitivity should be faded out to lateral directions of arrival. Section 5.2 examines this extension of the $\mathbf{r_E}$-vector model in more detail.

4.4.3 Optimal echo threshold slope τ

The generated data for $\varphi(\Delta t)$ can be used to find an optimal echo threshold slope parameter τ, which is used in eq. 8 to compute the time of arrival (TOA) dependent weight $w_{\tau,i}$ of the extended vector model from eq. 37. In order to find the optimal w_τ we define the optimization problem

$$\tau_{RMS} = \underset{\tau}{\text{argmin}} \left(J_{RMS}(\tau) \right). \tag{40}$$

The root mean sqare (RMS) cost function J_{RMS} is defined as

$$J_{RMS}(\tau) \;=\; \sqrt{\frac{1}{N \cdot M} \sum_{n=1}^{N} \sum_{m=1}^{M} |\Delta\varphi[n,m]|^2}. \tag{41}$$

Therefore the misalignment $\Delta\varphi[n,m]$ is defined as

$$\Delta\varphi[n,m] \;=\; \hat{\varphi}(\tau)[n,m] - \varphi[n,m], \tag{42}$$

with $\varphi[n]$ as the panning angle indicated by a subject in the listening test and $\hat{\varphi}(\tau)[n]$ as the panning angle which is computed with the extended $\mathbf{r_E}$-vector model (eq. 37) and

$$\hat{\varphi}(\tau) = \tan^{-1}\left(\frac{r_E[1](\tau)}{r_E[2](\tau)}\right), \quad \text{whereas} \quad \mathbf{r_E}(\tau) = \begin{pmatrix} r_E[1](\tau) \\ r_E[2](\tau) \\ r_E[3](\tau) \end{pmatrix}. \tag{43}$$

Index n indicates the panning angle indication for the m-th LS pair, Δt and stimulus for a subject. Because of the inconsistent indications for the lateral LS pairs this data is neglected. Also, panning angle indications whit an echo flag are neglected in this analysis. Furthermore, it is possible to merge the data for positive and negative ICTDs because of the symmetry of the frontal LS pairs. Consequently, accuracy is increased. Hence, there are N x M = 5 x 5 = 25 indications per stimulus and subject i. With eq. 41 and 42 it can be written

$$J_{RMS,i}(\tau) \;=\; \sqrt{\frac{1}{25} \sum_{n=1}^{5} \sum_{m=1}^{5} |\hat{\varphi}(\tau)[n,m] - \varphi_i[n,m]|^2}. \tag{44}$$

Figure 49 shows the cost functions $J_{RMS}(\tau)$ per subject for playback of PPN and clicks over the frontal LS pairs. First of all, we can see that there is a distinct minimum

nearly for all functions in the case with PPN. Consequently, an optimization would be beneficially. For playback of clicks this is not quite obvious. There are some $J_{RMS,i}(\tau)$ which have a distinct minimum but most curves tend to a $J_{RMS} \approx 10°$ for an decreasing τ.

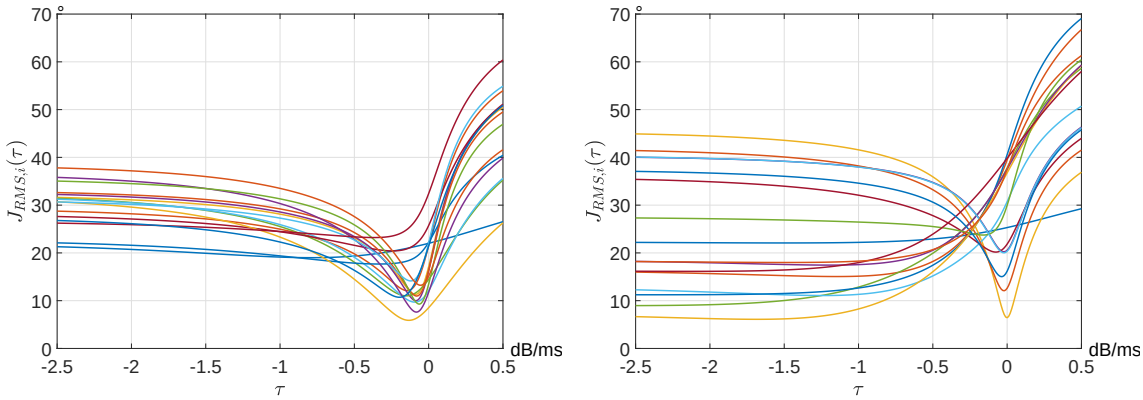


Figure 49 – Cost functions $J_{RMS}(\tau)$ per subject i for PPN (a) and clicks (b). Only data from frontal LS pairs is used.

A clustering is performed for $J_{RMS}(\tau)$ per subject i. For playback of PPN, only $J_{RMS}(\tau)$ from subject 1 has a less distinct minimum in comparison to these of the other subjects. Consequently, this $J_{RMS}(\tau)$ is neglected in further analysis. For playback of clicks there are two distinct clusters of $J_{RMS}(\tau)$. The first cluster consists of the curves from subjects: 1, 2, 5, 6, 8, 10, 11, 14 and 16. The second cluster holds $J_{RMS}(\tau)$ from the other subjects. These $J_{RMS}(\tau)$ have a more distinct minimum.

To merge the data of a cluster the median value $\overline{\varphi}[n]$ for every LS pair and ICTD Δt from direction indications of all subjects per cluster is computed. Eq. 44 can be rewritten as

$$J_{RMS}(\tau) \quad = \quad \sqrt{\frac{1}{25}\sum_{n=1}^{5}\sum_{m=1}^{5}|\hat{\varphi}(\tau)[n,m] - \overline{\varphi}[n,m]|^2}. \tag{45}$$

The minimum of $J_{RMS}(\tau)$ is found using the MATLAB function `fminunc`, which can be applied to unconstrained problems. Result of this minimum search as well as the behaviour of $J_{RMS}(\tau)$ are shown in figure 50. For PPN (fig. 50 (a)) there is a distinct minimum of $J_{RMS}(\tau)$ at τ_{opt} = -0.083 $\frac{\text{dB}}{\text{ms}}$. At this value for the echo threshold slope τ it is possible to reduce the residual RMS panning angle deviation about 1/2 to $J(\tau_{opt}) \approx 8.6°$ (cf. $J(0\frac{\text{dB}}{\text{ms}}) \approx 17°$). For playback of clicks the behaviour of $J_{RMS}(\tau)$ from the first cluster is nearly constant for $\tau \leq$ -1.5 $\frac{\text{dB}}{\text{ms}}$ (cf. fig. 50 (b)). However there is also an τ_{opt} = -2.083 $\frac{\text{dB}}{\text{ms}}$, which off course is not so obvious as in fig. 50 (a). The residual RMS panning angle deviation $J(\tau_{opt}) \approx 11.3°$ is reduced very well (cf. $J(0\frac{\text{dB}}{\text{ms}}) \approx 40°$). The resulting $J_{RMS}(\tau)$ for the second cluster has a distinct minimum for a $\tau_{opt} \approx 0$ $\frac{\text{dB}}{\text{ms}}$ (cf. fig. 50 (c)). This means that time-dependent weighting has nearly no influence

on the optimization results for the data of this cluster. Therefore, further analysis is focussed on the first cluster.

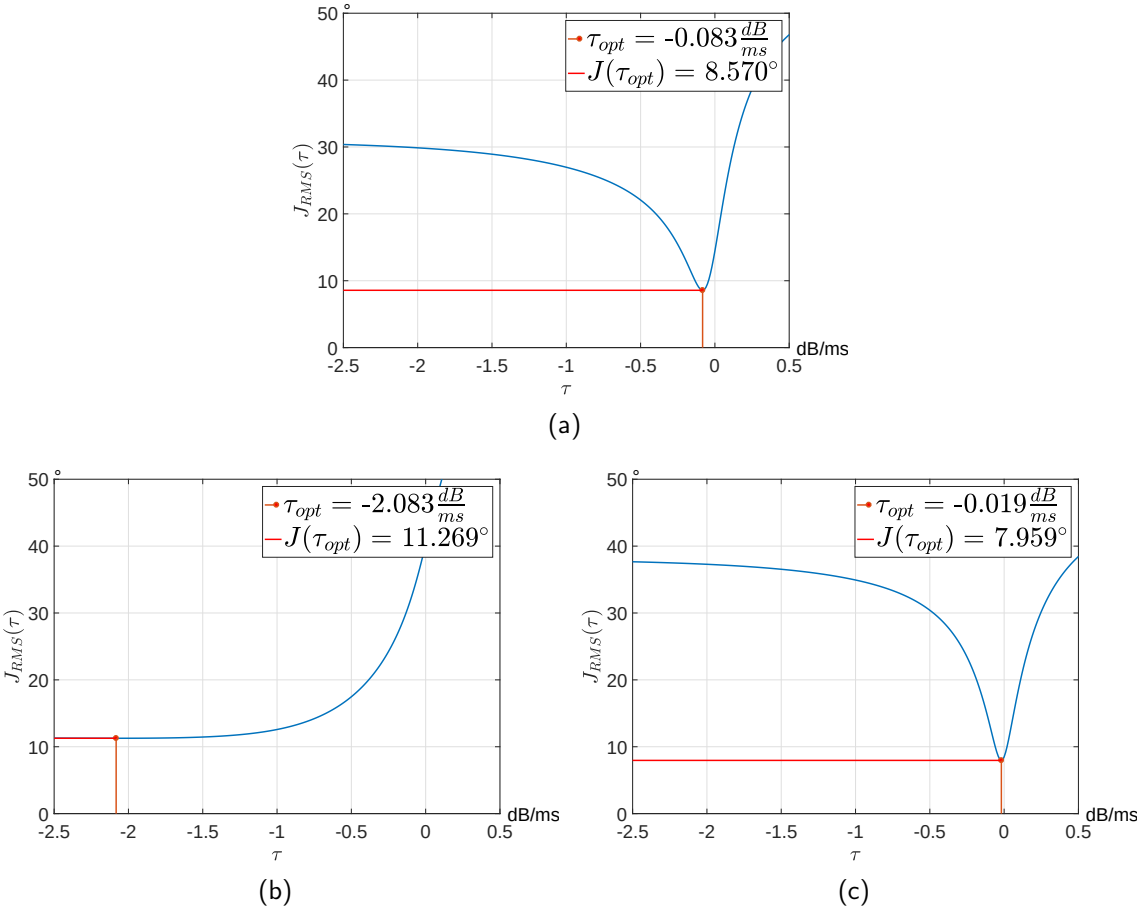


Figure 50 – Curve of $J_{RMS}(\tau)$, optimal echo threshold slope τ_{opt} and the residual relative angle RMS misalignment $J(\tau_{opt})$ for playback of PPN (a) and clicks for the first and second cluster ((b) and (c)).

Measurement data from the wide aperture LS pairs (6,17) and (5,18) has strong influences on the optimization problem. In order to reduce this influence a relative panning angle

$$\varphi_{rel}[n,m] = \frac{\varphi[n,m]}{\alpha_m} \tag{46}$$

is defined. α_m is the aperture angle for the m-th frontal LS pair according to tab. 4.

$$J_{RMS,rel}(\tau) = \sqrt{\frac{1}{25}\sum_{n=1}^{5}\sum_{m=1}^{5}\left|\frac{\hat{\varphi}(\tau)[n,m]-\overline{\varphi}[n,m]}{\alpha_m}\right|^2} \tag{47}$$

$$J_{RMS,rel}(\tau) = \sqrt{\frac{1}{25}\sum_{n=1}^{5}\sum_{m=1}^{5}|\Delta\varphi_{rel}[n,m]|^2} \tag{48}$$

Fig. 51 shows the behaviour of $J_{RMS,rel}(\tau)$, the optimal slope τ_{opt} and the residual relative angle RMS misalignment $J_{rel}(\tau_{opt})$ for playback of PPN (a) and clicks (b). Again both curves are different to each other. We can find a distinct minimum for PPN playback at τ_{opt} = -0.077 $\frac{\text{dB}}{\text{ms}}$ and the misalignment can be halved to $J_{rel}(\tau_{opt}) \approx 8\%$ (cf. $J_{rel}(0) \approx 13\%$). Also for the clicks, there is an optimal slope of τ_{opt} = -1.930 $\frac{\text{dB}}{\text{ms}}$ which is less distinct. However, the influence on the residual misalignment $J_{rel}(\tau_{opt}) \approx 10\%$ is much stronger (cf. $J_{rel}(0) \approx 40\%$). In comparison to the optimization with absolute panning angle differences $\hat{\varphi}[n,m]$ the optimal slope τ_{opt} becomes a little smaller.

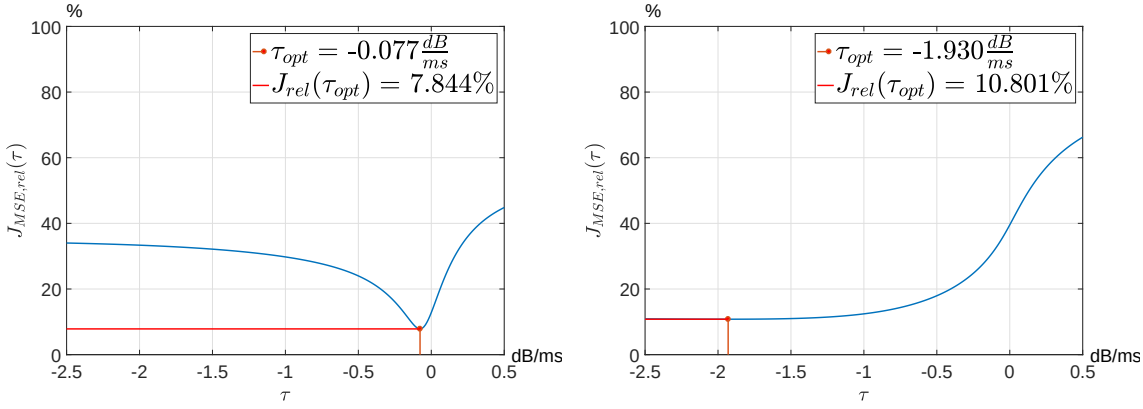


Figure 51 – Curve of $J_{RMS,rel}(\tau)$, optimal echo threshold slope τ_{opt} and the residual relative angle RMS misalignment $J_{rel}(\tau_{opt})$ for playback of PPN (a) and clicks (b).

In this section it could be shown that there is an optimal value for the parameter τ for the frontal playback results from the listening test. For more continuous signals, a $\tau_{PPN} \approx$ -0.1 $\frac{\text{dB}}{\text{ms}}$ and for more transient signals a $\tau_{clicks} \approx$ -1.9 $\frac{\text{dB}}{\text{ms}}$ was found. Consequently, it is reasonable to incorporate a signal dependent τ into the extended $\mathbf{r_E}$-vector model (see sec. 5.1).

4.4.4 ICTD panning curves

In the last step of data analysis from the listening experiment, we try to find ICTD-dependent panning curves. Because of the inconsistent data for lateral playback directions, we only use data of playback from the frontal LS pairs. Again to reduce the strong influence of the wide aperture LS pairs (6,17) and (5,18) we use the relative panning angle $\varphi_{rel}[n,m]$ according to eq. 46. If $\varphi_{rel}[n,m] = 0$, the virtual sound sources is perceived from the center position. If $\varphi_{rel}[n,m] = 1$, the subjects indicate the virtual sound source directly at the LS position. $\varphi_{rel}[n,m]$ is not dependent on the aperture angle α of the LS pairs. Consequently, it is possible to compute the median value $\overline{\varphi}_{rel}[n]$ and corresponding 95% confidence intervals for direction indications caused by the same Δt. In order to find a function for $\hat{\varphi}_{rel}(\Delta t)$ a nonlinear curve-fit in the least-squares

sense is performed according to

$$\hat{\varphi}_{rel}(\Delta t) = \frac{2}{\pi} \arctan\left(\psi \Delta t\right). \tag{49}$$

The $\arctan$-function seems to be appropriate because of its properties. For a $\Delta t \geq 0$ ms it is strictly monotonically increasing and it converges to 1 for $\Delta t \to \infty$. The parameter ψ determines the slope of $\hat{\varphi}_{rel}(\Delta t)$.

Figure 52 shows the median relative panning angles $\overline{\varphi}_{rel}$ with corresponding confidence intervals for PPN and clicks (blue). Between the $\overline{\varphi}_{rel}$ linear interpolation is performed. Both resulting curves have significantly different behaviour. For PPN, there is a nearly constant slope of $\frac{\overline{\varphi}_{rel}}{\Delta t} \approx 0.017\ \frac{1}{\text{ms}}$. Subjects perceived a maximal displacement of the virtual sound source of $\alpha/2$. For clicks already a small ICTD of $\Delta t = 5$ ms causes a large displacement of $\overline{\varphi}_{rel} \approx 0.9$, This is reached with a slope of $\frac{\overline{\varphi}_{rel}}{\Delta t} \approx 0.18\ \frac{1}{\text{ms}}$. Of course for a larger ICTD, this displacement remains nearly constant. There are no significant changes of $\overline{\varphi}_{rel}$ for $\Delta t > 5$ ms. Moreover, the resulting functions $\hat{\varphi}_{rel}(\Delta t)$ from the fitting according to eq. 49 are plotted. Again, the slope parameter ψ for PPN is approx. 10 times smaller than this for clicks. (cf. $\psi_{PPN} = 0.0441\ \frac{1}{\text{ms}}$ to $\psi_{clicks} = 0.5765\ \frac{1}{\text{ms}}$). Furthermore the echo detection threshold Δt_{echo} is plotted. This is done because for a $\Delta t > \Delta t_{echo}$ perception of an echo and in consequence source-splitting is very likely. Therefore at larger ICTDs, it is disadvantageous only to have a simple panning curve because two separate auditory events can be perceived. Summing localization is definitely no longer happening. For the playback, of PPN this is the case for $\Delta t > 28.78$ ms and for clicks $\hat{\varphi}_{rel}(\Delta t)$ for a $\Delta t > 19.95$ ms.

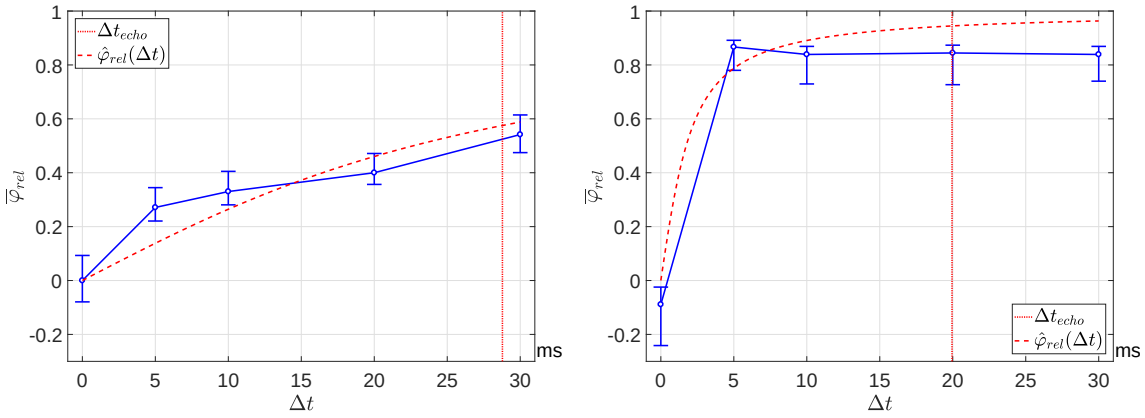


Figure 52 – Median relative panning angle $\overline{\varphi}_{rel}$ and fitted ICTD panning curve $\hat{\varphi}_{rel}(\Delta t)$ for (a): PPN ($\psi_{PPN} = 0.0441\ \frac{1}{\text{ms}}$) and (b): clicks ($\psi_{clicks} = 0.5765\ \frac{1}{\text{ms}}$). Moreover the echo detection threshold Δt_{echo} is plotted.

5 Further extensions

Based on the results from sec. 4.4.3 and sec. 4.4.2 we can integrate a signal-dependent echo threshold slope τ_{sig} and a fade-out function for the time-dependent weights $w_{\tau,i}$ for lateral playback directions into the extended energy vector model from eq. 37. Moreover, a qualitative assessment regarding the source-splitting effect at a certain ICTD Δt_{echo} is performed.

5.1 Signal-dependent slope τ_{sig}

To integrate the signal-dependent echo threshold parameter τ_{sig} into the $\mathbf{r_E}$-vector model, a simple blending according to eq. 50 is introduced. The parameter $0 \leq \beta \leq 1$ blends between τ_{PPN} and τ_{clicks}. For a scenario with more stationary sound playback ($\beta \to 0$) τ_{sig} approaches $\tau_{PPN} \approx$ -0.1 $\frac{\text{dB}}{\text{ms}}$. For a more transient sound ($\beta \to 0$) τ_{sig} becomes $\tau_{clicks} \approx$ -1.9 $\frac{\text{dB}}{\text{ms}}$.

$$\tau_{sig} = (1 - \beta) \cdot \tau_{PPN} + \beta \cdot \tau_{clicks} \tag{50}$$

Eq. 50 can be applied to eq. 8 from sec. 2.2 and the time-dependent weights $w_{\tau_{sig},i}$ can be computed according to

$$w_{\tau_{sig},i} = ((1 - \beta) \cdot \tau_{PPN} + \beta \cdot \tau_{clicks}) \cdot \Delta t_i. \tag{51}$$

Of course, $w_{\tau_{sig},i}$ can be deployed to the $\mathbf{r_E}$-vector model of eq. 37 from sec. 2.5.2. Figure 53 shows the mean localization error $\overline{\Delta\theta}$ in dependence on the listening position for an 8 LS array for slope blend parameter $\beta = 0$, 0.5, and 1. The simulation is done with 3rd-order Ambisonics, max-$\mathbf{r_E}$ weighting, without the image-source model and no gain and delay compensation. Radiation patterns are modelled for Genelec 8020 LSs for the frequency band above 5 kHz. The simulation clearly shows that for an increasing β the area with plausible localization shrinks dramatically. This result is comparable to human perception: More transient signals ($\beta \to 1$) can be localized much better than stationary ones. Consequently, $\overline{\Delta\theta}$ has to increase faster when moving outside the center position of a LS array.

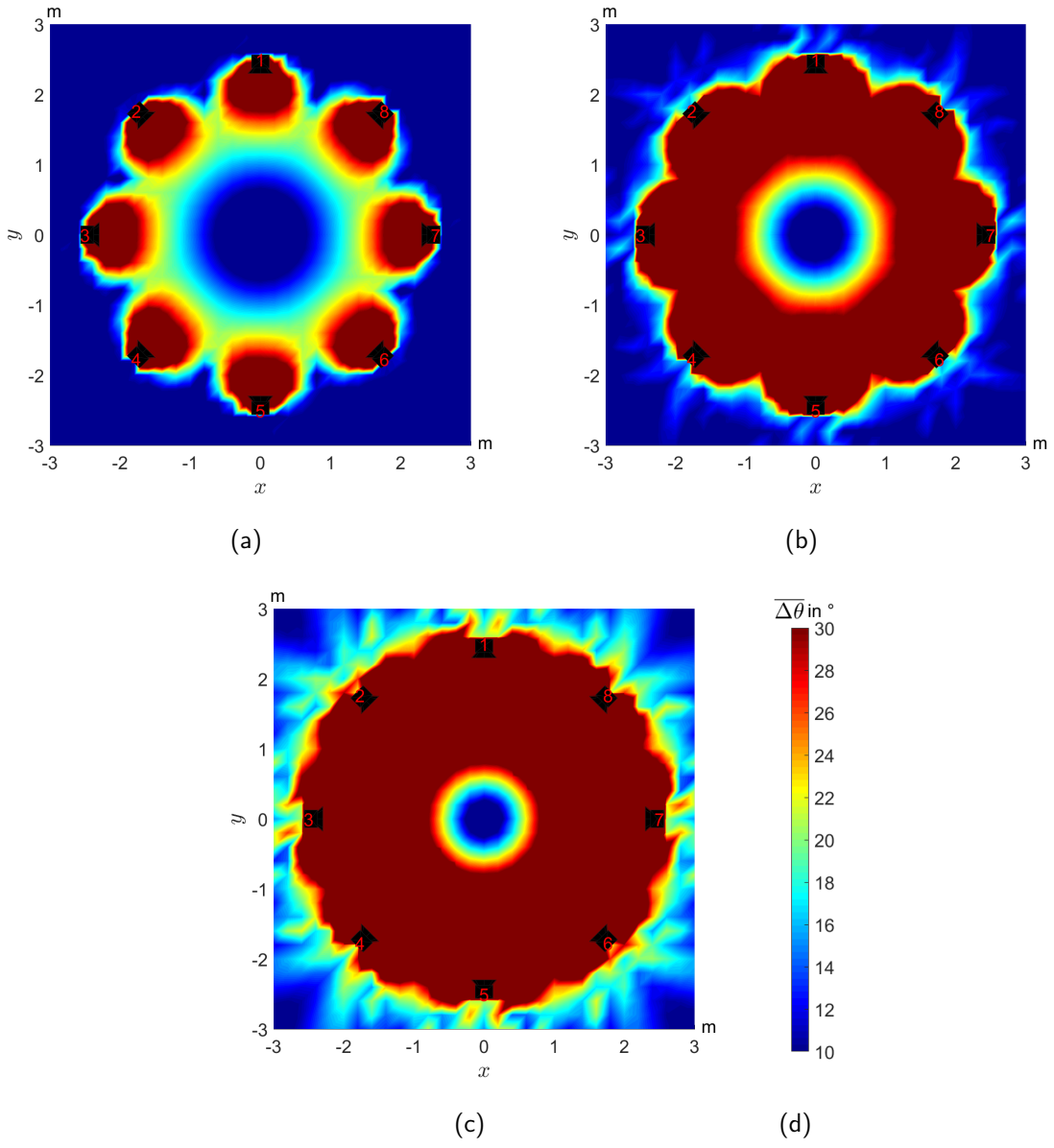


Figure 53 – Mean localization error $\overline{\Delta\theta}$ in dependence on the listening position for 3rd-order Ambisonics and max-$\mathbf{r_E}$ weighting for varying slope blend parameter β. No gain and delay compensation is done. Simulation is done without the image-source model. Radiation patterns are modelled for Genelec 8020 LSs for the frequency band above 5 kHz. Slope blend parameter is (a): $\beta = 0$, (b): $\beta = 0.5$, (c): $\beta = 1$.

5.2 Lateral fade-out of time-dependent weighting

As shown in sec. 4.4.1, panning angle discrimination based on ICTD for lateral sound sources is very inconsistent. Consequently, a blending function can be defined to influence damping weights $w_{d,i}$ and time-dependent weights $w_{\tau_{sig},i}$.

Because of common performance practice, the area around the center LS is chosen as main viewing direction (MVD) of the audience. The horizontal angle of the source-to-receiver-vector $\mathbf{r}_{\boldsymbol{\theta}_1,\mathbf{v}}$ for center LS (1) and a listening position $\mathbf{v}$ can be written as (cf. eq. 30 and 31)

$$\tilde{\varphi}_{\boldsymbol{\theta}_1,\mathbf{v}} = \tan^{-1}\left(\frac{r_{\boldsymbol{\theta}_1,\mathbf{v}}[1]}{r_{\boldsymbol{\theta}_1,\mathbf{v}}[2]}\right) \quad \text{with} \quad \mathbf{r}_{\boldsymbol{\theta}_1,\mathbf{v}} = \begin{pmatrix} r_{\boldsymbol{\theta}_1,\mathbf{v}}[1] \\ r_{\boldsymbol{\theta}_1,\mathbf{v}}[2] \\ r_{\boldsymbol{\theta}_1,\mathbf{v}}[3] \end{pmatrix}. \tag{52}$$

In order to incorporate the fact that not every listener is directly facing the center LS (1), the source-to-receiver-angle is limited to $|\tilde{\varphi}_{\boldsymbol{\theta}_1,\mathbf{v}}| \leq \tilde{\varphi}_{\boldsymbol{\theta}_1,\mathbf{v},max} \rightarrow \varphi_{MVD}$. Especially, this limitation is useful for listening positions in the lateral parts of the listening area. Furthermore, in this way the MVD is broadened to the area around the center LS (1). The blending angle χ is defined as the difference between the MVD angle and the source-to-receiver-angle of every LS of the array for a listening position

$$\chi = \tilde{\varphi}_{\boldsymbol{\theta}_i,\mathbf{v}} - \varphi_{MVD}. \tag{53}$$

Because of advantageous properties $\cos^2$- and $\sin^2$-terms are used for the blending function. Therefore, the lateral fade-out weight $w_{\chi,i}$ for every LS can be written as

$$w_{\chi,i} = \left(w_{\tau_{sig},i}\cos^2(\chi) + \sin^2(\chi)\right) \cdot w_{d,i}. \tag{54}$$

$w_{\chi,i}$ replaces damping weights $w_{d,i}$ and time-dependent weights $w_{\tau_{sig},i}$ in eq. 37. Consequently the $\mathbf{r}_\mathbf{E}$-vector model results in

$$\begin{aligned}\mathbf{r}_\mathbf{E} = &\frac{\sum_{i=1}^{I}\left(|w_{LS,i,j}(\hat{\varphi}_i,\hat{\theta}_i)w_{\chi,i}g_i|^{\overline{\gamma}_j}\boldsymbol{\theta}_i\right.}{\sum_{i=1}^{I}\left(|w_{LS,i,j}(\hat{\varphi}_i,\hat{\theta}_i)w_{\chi,i}g_i|^{\overline{\gamma}_j}\right.} \\ &\frac{\left.+\sum_{l=1}^{L}|(1-\overline{\alpha})^{K/2}w_{LS,i,l,j}(\hat{\varphi}_{i,l},\hat{\theta}_{i,l})w_{\chi,i,l}g_i|^{\overline{\gamma}_j}\boldsymbol{\theta}_{i,l}\right)}{\left.+\sum_{l=1}^{L}|(1-\overline{\alpha})^{K/2}w_{LS,i,l,j}(\hat{\varphi}_{i,l},\hat{\theta}_{i,l})w_{\chi,i,l}g_i|^{\overline{\gamma}_j}\right)}.\end{aligned} \tag{55}$$

Figure 54 shows $\overline{\Delta\theta}$ for the 8 LS array for a varying maximum source-to-receiver-angle $\varphi_{u_1,v,max}$. Simulation conditions are similar to those from fig. 53. The slope blend parameter is set to $\beta = 0.639$ and the MVD is set to LS (1). In fig. 54 (a) the maximum source-to-receiver-angle is set to $\varphi_{u_1,v,max} = 0°$. This means that listeners at all listening positions would face along the positive y-axis. In comparison to fig. 53 (c) the listening area for plausible localization is enhanced significantly but the minimum $\overline{\Delta\theta}$ in the center

of the array is larger. The listening area has the shape of a squashed hexagon. For a $\varphi_{u_1,v,max} = 20°$ resp. 50° ((b) and c(c)) the hexagon is further squashed in the rear parts and slightly stretched in the frontal lateral parts. This effect is caused through different $\varphi_{u_1,v}$ for the front and the rear. However the size of the listening area does not change dramatically. The differences between fig. 53 (b) and (c) are marginal. A larger $\varphi_{u_1,v,max}$ seems to have less influence on the simulation results.

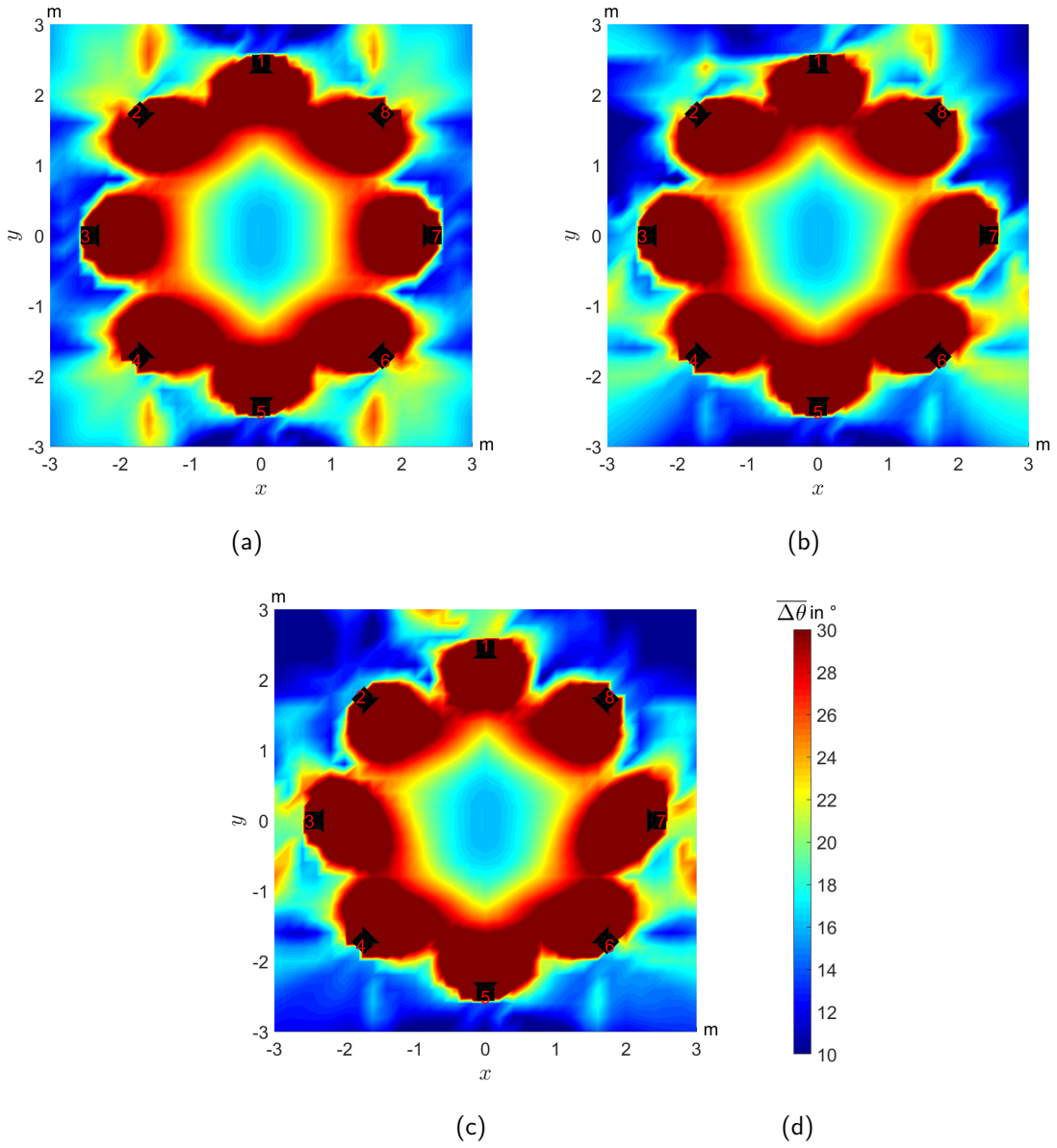


Figure 54 – Mean localization error $\overline{\Delta\theta}$ in dependence on the listening position for 3rd-order Ambisonics, max-$\mathbf{r_E}$ weighting and slope blend parameter $\beta = 0.639$ for varying $\varphi_{u_1,v,max}$. No gain and delay compensation is done. Simulation is done without the image-source model. Radiation patterns are modelled for Genelec 8020 LSs for the frequency band above 5 kHz. The maximum source-to-receiver-angle for LS (1) is (a): $\varphi_{u_1,v,max} = 0°$, (b): $\varphi_{u_1,v,max} = 20°$, (c): $\varphi_{u_1,v,max} = 50°$.

5.3 Source-splitting ICTD Δt_{echo}

From the investigations in sec. 4.4.2 we get the two ICTDs $\Delta t_{echo,PPN} \approx 29$ ms and $\Delta t_{echo,clicks} \approx 20$ ms at which echo detection resp. source-splitting with a probability of 50% is likely. Of course, this threshold values were determined only for frontal playback situations and without ICLDs. In the following we want to look in more detail into the vectorial representation of a possible source-splitting case by the example plotted in fig. 55. In this figure the dominance-vectors per LS as well as the resulting $\mathbf{r_E}$-vector are plotted for playback with LS pair (1,8) of the IEM CUBE. The ICLD of the LS pair is chosen with respect to a common playback situation with ΔL = -12 dB. The vectors are computed at four points P_1 to P_4 with an angle $\theta = 160°$ and radii of $r_1 = 1.3$ m, $r_2 = 1.6$ m, $r_3 = 1.9$ m and $r_4 = 2.2$ m. The dominance-vectors are computed regarding to eq. 56. The $\mathbf{r_E}$-vectors are computed with respect to eq. 37. The vectors are scaled differently for a better illustration. Further parameters of the simulation can be found in the caption.

$$\mathbf{s}_i = |w_{LS,i,j}(\hat{\varphi}_i, \hat{\theta}_i) w_{\tau_{sig},i} w_{d,i} g_i|^{\overline{\gamma}_j} \boldsymbol{\theta}_i \tag{56}$$

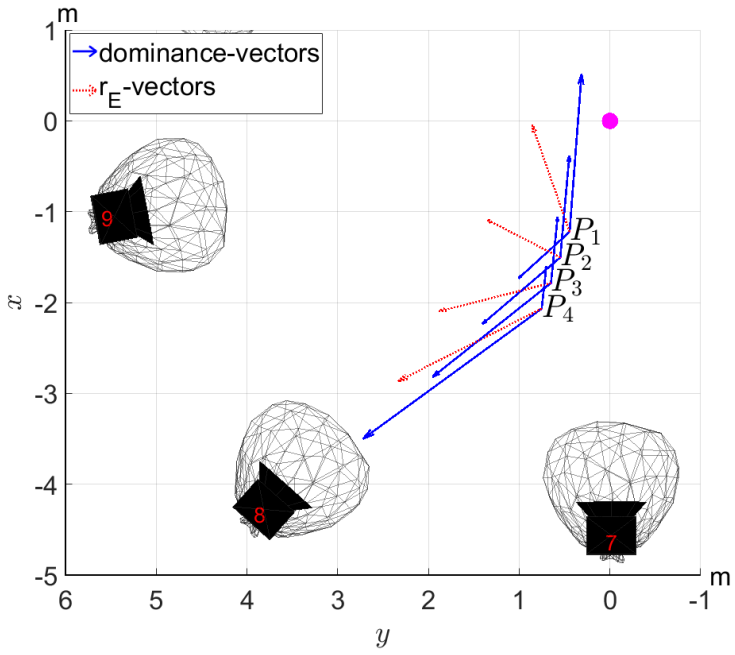


Figure 55 – Source-splitting case study for playback with LS pair (1,8) of the IEM CUBE. Dominance-vectors and $\mathbf{r_E}$-vectors at points P_1 to P_4. Simulation with ΔL = -12 dB, no delay compensation, without an image-source model, $\beta = 0.639$ and without a lateral fade-out of time-dependent weights $w_{\tau_{sig},i}$. Radiation patterns are modelled for Tannoy 1200 LSs for the frequency band above 5 kHz.

Furthermore, the impulse responses with their corresponding echo threshold functions for points P_1 to P_4 and the echo detection threshold $\Delta t_{echo,clicks}$ are plotted in fig. 56. The following computations are based on investigations in [43]. The echo threshold functions can be computed with

$$F_{echo}(t) = \max(g(T_i) * f_{echo}(t)) \quad \text{with} \quad g(T_i) = g_i. \tag{57}$$

Thereby, the prototype of the echo threshold function is

$$f_{echo}(t) = \begin{cases} 60\text{dB/ms} & , \quad 0 \leq t \leq 1\text{ms} \\ \tau_{clicks} & , \quad t \geq 1\text{ms} \end{cases}. \tag{58}$$

f_{echo} can be interpreted as a time-dependent masking function. Through convolution of f_{echo} with a single impulse g_{T_i} we get a time-shifted and scaled version of f_{echo} per impulse. This shifted echo threshold functions can be superimposed and we get a resulting function $F_{echo}(t)$. Let us assume that delayed arriving signals were masked by $F_{echo}(t)$. Moreover, source splitting only occurs when a delayed incoming signal with a minimum $\Delta t \geq \Delta t_{echo}$ is no longer masked by $F_{echo}(t)$.

The listening positions P_1 to P_4 of fig. 55 and 56 are chosen with respect to four different playback situations regarding source-splitting. First of all, we can see that, regarding $\Delta t_{echo,clicks}$ and the definition from above, none of these cases causes the source-splitting effect. However, from fig. 48 (b) we have a function for $\widehat{EDR}(\Delta t)_{clicks}$ which gives a echo detection probability for a certain Δt. With eq. 39 and the parameters of tab. 7 the equation for this function is

$$\widehat{EDR}(\Delta t)_{clicks} \quad = \quad 0.325 \cdot \ln(0.17 \Delta t \frac{1}{\text{ms}} + 1.264). \tag{59}$$

Unfortunately, this echo detection probability is for a ICLD $= 0$ dB. When moving from listening position P_1 to P_4 the signal from LS (1) $\rightarrow s_{LS1}$ is decreasing and arriving with increasing delay, whereas the signal from LS (8) $\rightarrow s_{LS8}$ is increasing and arriving with decreasing delay. The objective is to find the listening position where source-splitting is most likely.

The four cases are:

- $\mathbf{P_1}$ (fig. 56 (a)):
  At this listening position s_{LS1} is not masked by s_{LS8}.
  The $\widehat{EDR}(3.6\text{ms})_{clicks} \approx 20\%$ is relatively small and would be even more smaller if the large level of s_{LS1} would be considered. Also the resulting $\mathbf{r_E}$-vector indicates that source-splitting is unlikely.
- $\mathbf{P_2}$ (fig. 56 (b)):
  Again, s_{LS1} is not masked by s_{LS8}. Both signals have nearly the same level and the echo detection becomes more likely $\widehat{EDR}(5.2\text{ms})_{clicks} \approx 25\%$. Also the $\mathbf{r_E}$-vector indicates a direction in between both LS directions.
- $\mathbf{P_3}$ (fig. 56 (c)):
  Comparatively, this case has a large $\widehat{EDR}(6.8\text{ms})_{clicks} \approx 29\%$. The level of s_{LS1} is barely large enough to be not masked by s_{LS8}. Because of the large level of s_{LS8} source-splitting would be even more likely.

- $\mathbf{P_4}$ (fig. 56 (d)):
  Now s_{LS1} is masked by s_{LS8}. Source-splitting is very unlikely. Also the resulting $\mathbf{r_E}$-vector is points towards LS (8). ($\widehat{EDR}(8.5\text{ms})_{clicks} \approx 32\%$).

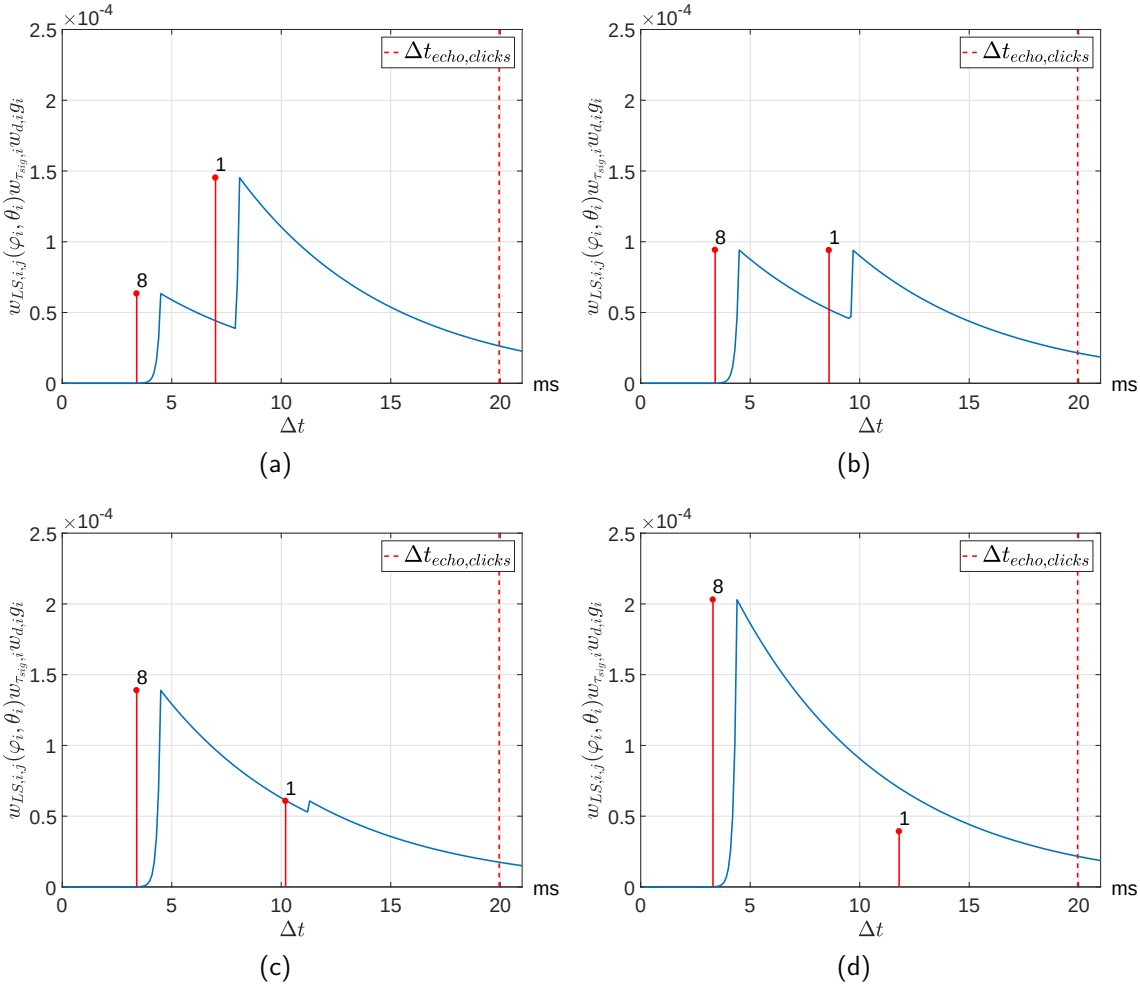


Figure 56 – Impulse responses with their corresponding echo threshold functions for points P_1 to P_4 ((a) to (d)) for simulation setup from fig. 55. Furthermore, echo detection threshold $\Delta t_{echo,clicks}$ is plotted.

6 Conclusion

This master thesis investigated several extensions of the energy vector ($\mathbf{r_E}$-vector). It is used as a simple but efficient tool for prediction of the listening area with plausible localization inside spherical LS arrays. For this purpose an already existing extended $\mathbf{r_E}$-vector model that was suggested in [43] by the author himself was the basis for further development. This vector model integrates sound dissipation and time-dependency of the incoming sound signals at a receiving point into the basic vector model from [33].

The first extension done in this master thesis incorporated a frequency-dependent panning according to investigations from Helm and Kurz [37]. Therefore a frequency-dependent slope parameter $\gamma(f)$ is used as exponent in the vector equation 13. For reasons of computational effort there are three frequency bands for $\overline{\gamma}_j$: 100Hz $\leq f \leq$ 1kHz, 1kHz $\leq f \leq$ 5kHz, and $f \geq$ 5kHz. Next, a simple n-th order three-dimensional image-source model was integrated to investigate influences from reflections at boundary surfaces of the room. Moreover, sound absorption over a mean absorption coefficient $\overline{\alpha}$ is included in the image-source model. If the mean absorption coefficient for the room is not known, an optimal $\overline{\alpha}$ can be computed from the reverberation time equations by Sabine or Eyring. Furthermore, sum equations for the number of mirror sources could be found for 2-D and 3-D models (see Appendix). Also LS radiation patterns were integrated into the $\mathbf{r_E}$-vector model. Initial measurements of three-dimensional impulse responses of existing LS were done with a circular microphone array in the anechoic measurement chamber of the IEM. The measurements were done referring to work from Brandner et al. [5]. The impulse responses were transformed into the frequency domain and a three-octave-smoothing is done. Mean gains from this smoothed magnitude spectra are computed according to the frequency bands from the frequency-dependent slope parameter $\overline{\gamma}_j$. The phase of the impulse responses is neglected. For compact representation of the radiation patterns, the SH domain was used. To find the wave spectra $\Psi_{n,m,j}$, a optimization with the mean gains was performed. For illustration of the LS radiation patterns, the SH are evaluated at the points of a t-design. By defining a source-to-receiver-vector $\mathbf{r}_{\boldsymbol{\theta}_\mathbf{i},\mathbf{v}}$ the weights can be computed for every LS in every listening position. This weights can be integrated into the vector model (cf. 37).

Another objective of this master thesis was the investigation of the echo threshold slope τ that mainly influences the time-dependent weights $w_{\tau,i}$ of the vector model. At first a literature study was performed to get an overview of already existing research regarding this topic. It could be found out that for stereophonic setups, which are building the basis for spherical LS arrays, there are in general three different cases for localization of an auditory event depending on the ICTD. For an ICTD of 0 to 1 ms summing localization is dominant and one distinct phantom sound source is perceptible. Investigations regarding summing localization mainly focus on equivalence stereophony and trading curves. The precedence effect is present for ICTDs of 1 to 5 ms for clicks and up to $\approx$ 50 ms for speech. The auditory event is perceived from the leading LS. In this interval, several effects like comb filters or source widening are perceptible and the phantom sound source splits continuously into two seperate auditory events. For higher ICTDs the signal at the lagging LS is clearly perceived as an echo. Lead-lag-experiments are mainly used to

investigate the precedence effect and the echo threshold. Experiments from the studied literature resulted in echo threshold slopes of -0.19 $\frac{\text{dB}}{\text{ms}} \leq \tau \leq$ -0.84 $\frac{\text{dB}}{\text{ms}}$ depending on the spectrum, the temporal envelope, the loudness and the impulse density of the used stimuli as well as direction and aperture angle of the used LS pair.

In order to find an optimal τ for the $\mathbf{r_E}$-vector model, a listening test was designed and conducted. For the listening experiment a circular array with 21 LS was set up in the IEM CUBE. In this adjustment experiment panning only was done via ICTDs. Playback was realized with frontal and lateral LS pairs with various aperture angles α. A moveable sound source should be panned to the direction of the perceived phantom sound source. Panning was realized with VBAP. The adjusted panning angle was used for further analysis. Moreover, it was possible to indicate the perception of an echo. The ICTDs $\Delta t = 0$, $\pm$5, $\pm$10, $\pm$20, $\pm$30 ms played back over 5 frontal and 4 lateral LS pairs. Pulsed pink noise and clicks were used as stimulus signals. 16 subjects participated the listening experiment. It could be shown that localization based on ICTD is possible for frontal playback directions and lateralization is more distinct for more transient signals. Moreover, for an increasing aperture angle α and an increasing Δt the localization accuracy decreases. For lateral playback directions, consistent localization based on ICTD is barely possible. Subjects tend to localize the phantom sound source mainly from the front, also for a Δt that should cause a localization to the back. For an increasing α, the inconsistency in localization also is increasing. These results are mainly caused by the fact that the lateral LS pair are placed on the cone-of-confusion and therefore front/back-confusion is triggered. The data from the echo detection was used to find the echo threshold Δt_{echo}. It could be shown that more a transient stimulus signal resp. larger aperture angle α cause a larger $\overline{EDR}$ and a smaller Δt_{echo}. Unfortunately, $\Delta t_{echo,PPN} \approx$ 29 ms and $\Delta t_{echo,clicks} \approx$ 20 ms only could be determined for frontal playback, because echo discrimination for lateral playback is very inconsistent. In order to find an optimal echo threshold slope, an optimization of τ from the $\mathbf{r_E}$-vector model from eq. 37 was performed to match the listening test data. For both stimulus signals, a distinct minimum could be found ($\tau_{PPN} \approx$ -$\frac{1}{8}$ $\frac{\text{dB}}{\text{ms}}$ and $\tau_{clicks} \approx$ -$\frac{5}{4}$ $\frac{\text{dB}}{\text{ms}}$). Consequently, an integration of a signal-dependent τ into the extended vector model is reasonable. Finally, ICTD panning curves based on the relative panning angle indications φ_{rel} were determined for PPN and clicks. At playback of PPN $\overline{\varphi_{rel}}$ is increasing nearly with constant slope for an increasing Δt. However for clicks $\overline{\varphi_{rel}}$ is increasing very strong for 0 ms $\leq \Delta t \leq$ 5 ms and remains nearly constant at large a lateralization for larger Δt.

Based on the results of the listening experiment, a signal-dependent echo threshold slope τ_{sig} as well as a fade-out for the time-dependent weights $w_{\tau,i}$ for lateral playback directions was implemented. τ_{sig} is computed with the blending parameter β, which blends between τ_{PPN} ($\beta = 0$) and τ_{clicks} ($\beta = 1$). Simulations showed very well that an increasing β leads to a decreasing listening area. For the calculation of the fade-out of $w_{\tau,i}$ a main viewing direction, facing the center LS of an array, is assumed. A combined weight $w_{\chi,i}$ is defined, which blends out $w_{tau_{sig},i} w_{d,i}$ with a $\cos^2$-function and blends in $w_{d,i}$ with a $\sin^2$-function. Both functions are depending on the blending angle χ (cf. 53). Finally, a qualitative assessment regarding the source-splitting effect in dependence on Δt_{echo} is performed to find a listening position where source-splitting is most likely.

7 Further investigations

The results of the performed listening experiment show that a further investigation of the precedence effect, the echo threshold Δt_{echo} and the echo threshold slope τ has to be done. Therefore, listening experiments should be performed. These listening experiments should investigate other lateral and elevated LS positions as well as playback from three or more LS. Furthermore, a finer resolution of the ICTD steps would be desirable to find more reliable echo thresholds Δt_{echo}. Also, the dependence of τ on the ICLD ΔL should be investigated.

Further evaluation of the derived $\mathbf{r_E}$-vector model (eq. 55) is reasonable. Therefore other spherical arrays could be investigated. In Graz there are those from the Györgi-Ligeti-Saal in the MUMUTH of the University of Music and Performing Arts or the mAmbA from IEM [27]. However, also compact spherical LS arrays like the IEM IKO [82] or the IEM cubes [13] can be used for evaluation. Figure 57 shows results from an already existing listening experiment conducted by Wendt in 2018 [77]. The plotted localization directions (coloured vectors from the listening position to the localization ellipses) are computed with the extended $\mathbf{r_E}$-vector model.

Moreover the integration of the existing simulation with extended $\mathbf{r_E}$-vector model into the AllRADecoder of the IEM Plug-in Suite [9] would be quite useful to provide the user a tool for estimating the listening area for the desired LS array.

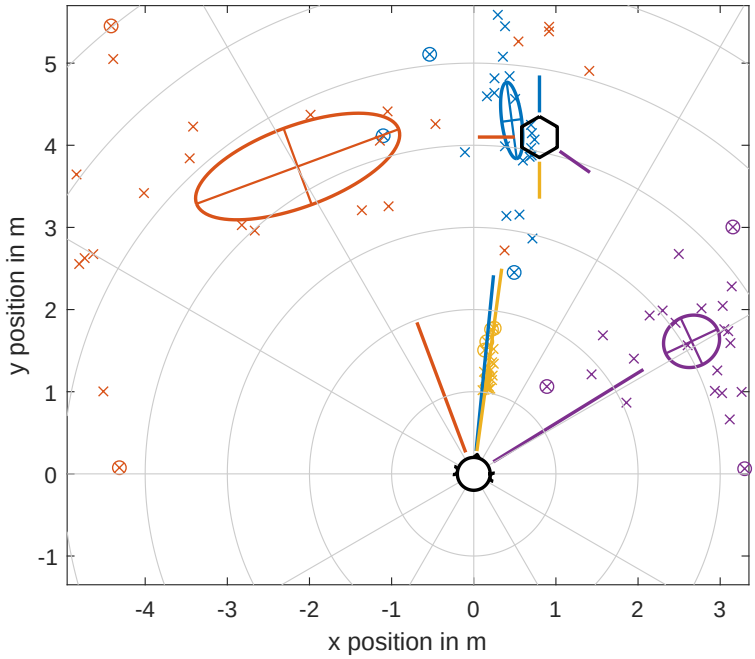


Figure 57 – Results from investigations regarding the localization at playback with the IEM IKO [77].

9. `https://plugins.iem.at`

8 Appendix

Proof for the number of mirror sources by induction.

• Proof for 2-D image-source model that:

$$\boxed{\sum_{n=1}^{N} 4n = 2N^2 + 2N} \tag{60}$$

$N = 1$:

$$4 \cdot 1 = 5 = 2 \cdot 1^2 + 2 \cdot 1$$

$N \to N + 1$:

$$\begin{aligned}
\sum_{n=1}^{N+1} 4n &\overset{!}{=} 2(N+1)^2 + 2(N+1) \\
\sum_{n=1}^{N+1} 4n &= \sum_{n=1}^{N} 4n + 4(N+1) \\
&= 2N^2 + 6N + 4 \\
&= 2(N^2 + 2N + 1) + 2(N+1) \\
&= 2(N+1)^2 + 2(N+1) \quad q.e.d.
\end{aligned}$$

• Proof for 3-D image-source model that:

$$\boxed{\sum_{n=1}^{N} 4n^2 + 2 = \frac{4}{3}N^3 + 2N^2 + \frac{8}{3}N} \tag{61}$$

$N = 1$:

$$4 \cdot 1^2 + 2 = 7 = \frac{4}{3} \cdot 1^3 + 2 \cdot 1^2 + \frac{8}{3} \cdot 1$$

$N \to N + 1$:

$$\begin{aligned}
\sum_{n=1}^{N+1} 4n^2 + 2 &\overset{!}{=} \frac{4}{3}(N+1)^3 + 2(N+1)^2 + \frac{8}{3}(N+1) \\
\sum_{n=1}^{N+1} 4n^2 + 2 &= \sum_{n=1}^{N} 4n^2 + 2 + 4(N+1)^2 + 2 \\
&= \frac{4}{3}N^3 + 2N^2 + \frac{8}{3}N + 4(N+1)^2 + 2 \\
&= \frac{4}{3}N^3 + 6N^2 + \frac{32}{3}N + 6 \\
&= \underbrace{\frac{4}{3}N^3 + 4N^2 + 4N + \frac{4}{3}} + \underbrace{2N^2 + 4N + 2} + \underbrace{\frac{8}{3}N + \frac{8}{3}} \\
&= \frac{4}{3}(N^3 + 3N^2 + 3N + 1) + 2(N^2 + 2N + 1) + \frac{8}{3}(N+1) \\
&= \frac{4}{3}(N+1)^3 + 2(N+1)^2 + \frac{8}{3}(N+1) \quad q.e.d
\end{aligned}$$

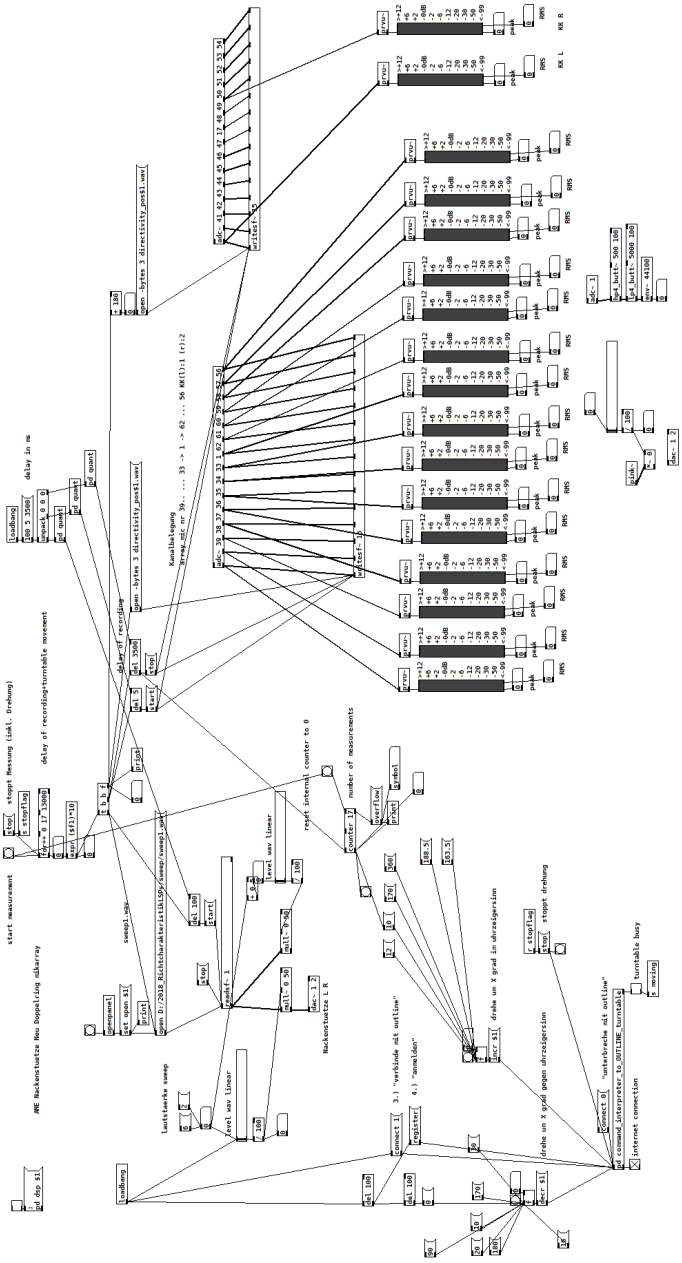


Figure 58 – Pd patch for the measurement of the LS radiation patterns.

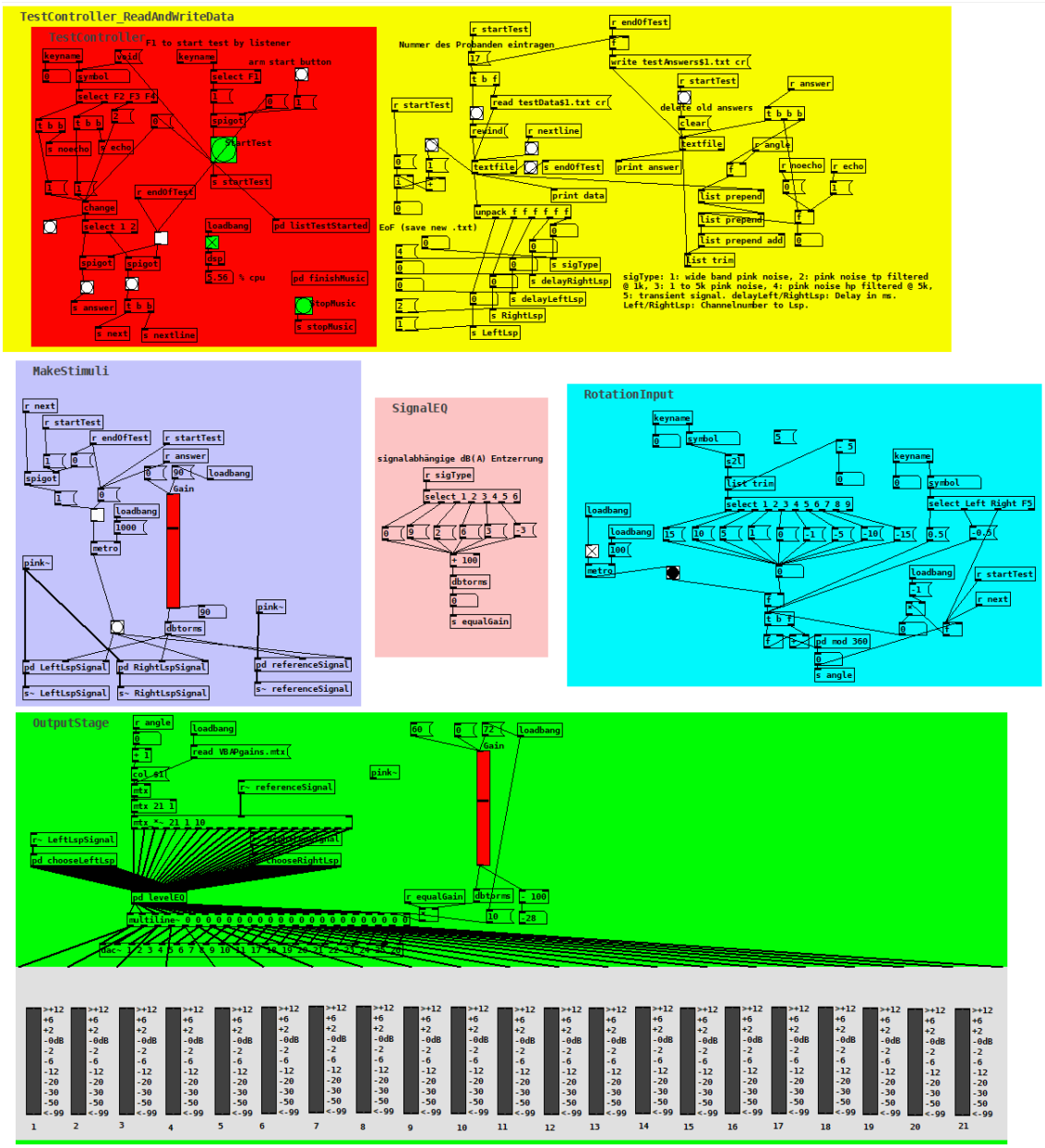


Figure 59 – Pd patch for the listening experiment.

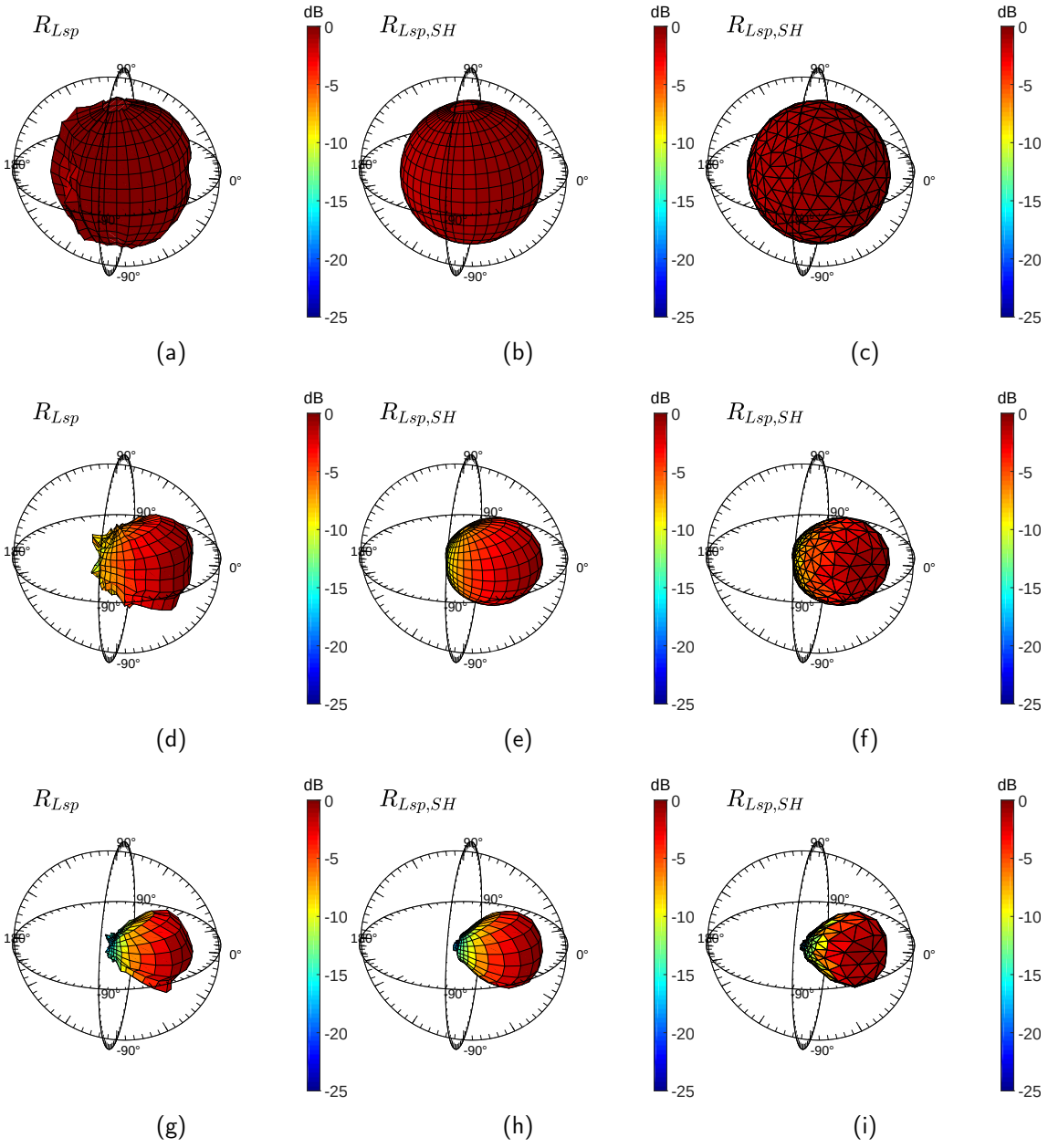


Figure 60 – Radiation patterns for a Lambda Labs CX1A LS for the frequency band from 100Hz to 1kHz (a,b,c), 1kHz to 5kHz (d,e,f) and above 5kHz (g,h,i). (a,d,g): Radiation pattern from measurement. (b,e,h): Radiation pattern in 1st/3rd/5th-order spherical harmonics. (c): Radiation pattern in 1st/3rd/5th-order spherical harmonics sampled on a 21st t-design.

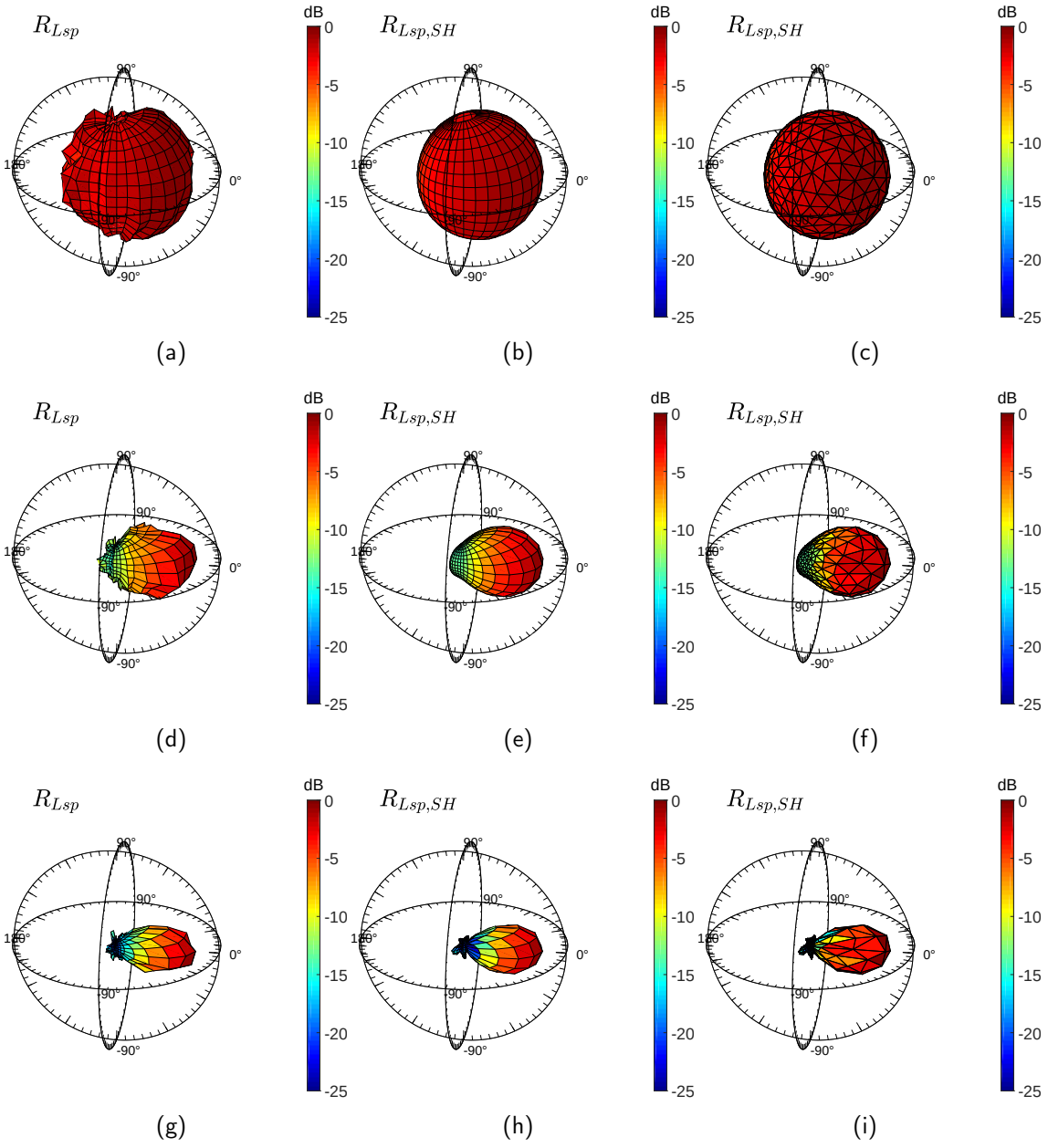


Figure 61 – Radiation patterns for a Tannoy 1200 LS for the frequency band from 100Hz to 1kHz (a,b,c), 1kHz to 5kHz (d,e,f) and above 5kHz (g,h,i). (a,d,g): Radiation pattern from measurement. (b,e,h): Radiation pattern in 1st/3rd/5th-order spherical harmonics. (c): Radiation pattern in 1st/3rd/5th-order spherical harmonics sampled on a 21st t-design.

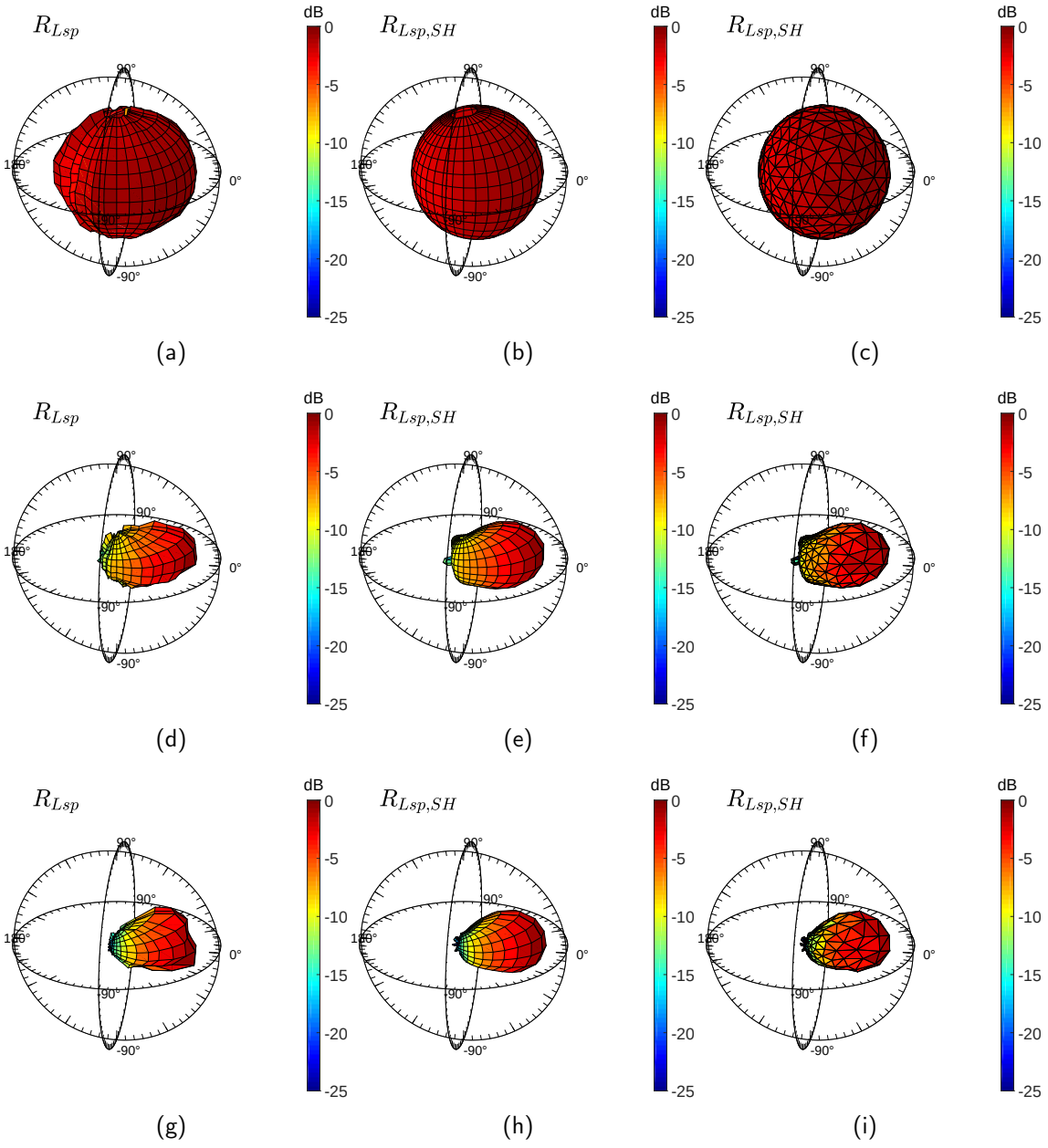


Figure 62 – Radiation patterns for a Neumann KH120A LS for the frequency band from 100Hz to 1kHz (a,b,c), 1kHz to 5kHz (d,e,f) and above 5kHz (g,h,i). (a,d,g): Radiation pattern from measurement. (b,e,h): Radiation pattern in 1st/3rd/5th-order spherical harmonics. (c): Radiation pattern in 1st/3rd/5th-order spherical harmonics sampled on a 21st t-design.

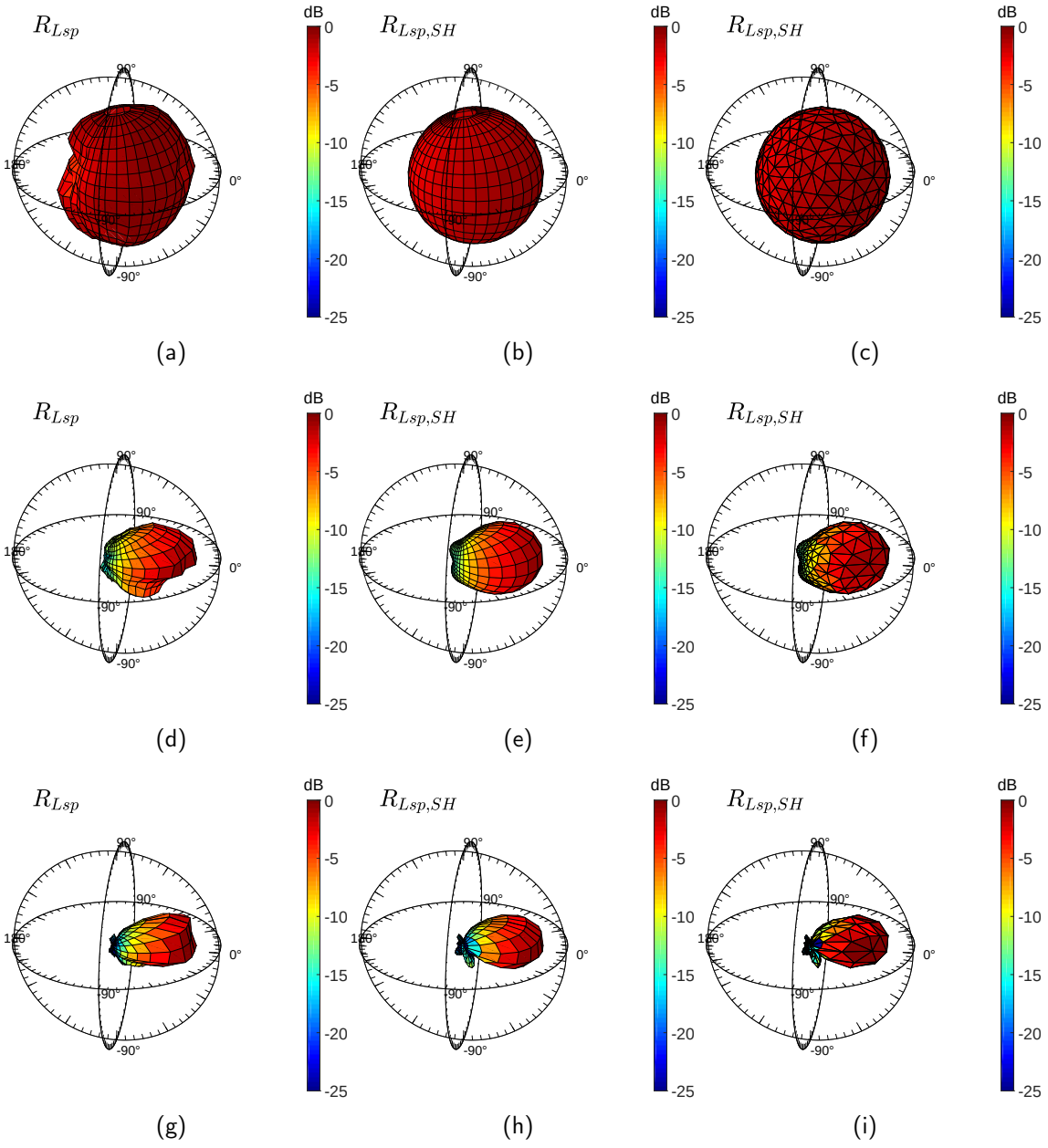


Figure 63 – Radiation patterns for a Yamaha DXR8 LS for the frequency band from 100Hz to 1kHz (a,b,c), 1kHz to 5kHz (d,e,f) and above 5kHz (g,h,i). (a,d,g): Radiation pattern from measurement. (b,e,h): Radiation pattern in 1st/3rd/5th-order spherical harmonics. (c): Radiation pattern in 1st/3rd/5th-order spherical harmonics sampled on a 21st t-design.

9 References

[1] A. J. Berkhout: „A Holographic Approach to Acoustic Control“, *Journal of the Audio Engineering Society*, vol. 36, pp. 977-995, December 1988

[2] A. J. Berkhout, D. de Vries and P. Vogel: „Acoustic control by wave field synthesis“, *Journal of the Acoustical Society of America*, vol. 93, pp. 2764-2778, 1993

[3] A. J. Berkhout, M. M. Boone, D. de Vries and P. Vogel: „Generation of sound fields using wave field synthesis, an overview“, *Proceedings of Active 95*, 1995

[4] J. Blauert: „Räumliches Hören“, *Monographien der Nachrichtentechnik*, S. Hirzel Verlag, Stuttgart, 1974

[5] M. Brandner, M. Frank and D. Rudrich: „DirPat - Database and Viewer of 2D/3D Directivity Patterns of Sound Sources and Receivers“, *144th AES Convention*, e-Brief 425, Milan, Italy, May 2018

[6] W. Burgtorf: „Untersuchungen zur Wahrnehmbarkeit verzögerter Schallsignale“, *Acustica 11*, pp. 97 - 111

[7] D. H. Cooper and T. Shiga: „Discrete-matrix multichannel stereo“, *J. Audio Eng. Soc.*, vol. 20, no. 5, pp. 346-360, 1972

[8] P. Craven and M. A. Gerzon: „Coincident microphone simulation covering three dimensional space and yielding various directional outputs“, *U.S. Patent*, no. 4,042,779, 1977

[9] L. Cremer: „Die wissenschaftlichen Grundlagen der Raumakustik“, Band 1, Stuttgart, S. Hirzel Verlag, 1948

[10] P. Damaske: „Die psychologische Auswertung akustischer Phänomene“, *7th International Congress on Acoustics*, Budapest, 21 G 2

[11] J. Daniel: „Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia“, *Ph.D. Dissertation*, University of Paris, 2000

[12] L. Davis and H. Lee: „Echo Thresholds for a 3D Loudspeaker Configuration“, *140th AES Convention*, e-Brief 239, Paris, France, June 2016

[13] T. Deppisch, N. Meyer-Kahlen, F. Zotter and M. Frank: „Surround with Depth on First-Order Beam-Controlling Loudspeakers“, *144th AES Convention*, Milan, Italy, 2018

[14] K. de Boer:„Stereofonische Geluidsweergave“, *Ph.D. Dissertation*, Technical University Delft, 1940

[15] M. Dietz, S. D. Ewert and V. Hohmann: „Auditory model based direction estimation of concurrent speakers from binaural signals“, *Speech Communication*, vol. 53, no. 5, pp. 592-605, May 2011

[16] „EBU SQAM CD: Sound Quality Assessment Material recordings for subjective tests“, 2008

[17] P. Felgett: „Ambisonic reproduction of directionality in surround-sound systems“, *Nature*, vil. 252, pp. 534-538, 1974

[18] N. V. Franssen: „Some considerations on the mechanism of directional hearing“, *Ph.D. Dissertation*, Technical University Maastricht, 1961

[19] N. V. Franssen: „Stereophony“, *Philips Technical Bibliography*, Eindhoven, 1963

[20] M. Frank, F. Zotter and A. Sontacchi: „Localization Experiments Using Different 2D Ambisonics Decoders“, *25th Tonmeistertagung*,Leipzig, Deutschland, November 2008

[21] M. Frank and F. Zotter: „All-Round Ambisonic Panning and Decoding“, *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 807-820, Oktober 2012

[22] M. Frank: „Phantom Sources using Multiple Loudspeakers in the Horizontal Plane“, *Ph.D. Dissertation*, Universität für Musik und darstellende Kunst Graz, Juni 2013

[23] M. Frank: „Source Width of Frontal Phantom Sources: Perception, Measurement, and Modeling“, *Archives of Acoustics*, vol. 38, no. 3, pp. 311-319, Januar 2013

[24] M. Frank: „Localization Using Different Amplitude-panning Methods in the Frontal Horizontal Plane“, *Proc. of EAA Joint Symposium on Auralization and Ambisonics, 3-5 April*, Berlin, pp. 41-47, April 2014

[25] M. Frank: „Simple Uncertainty Prediction for Phantom Source Localization“, *DAGA Tagungsbericht*, Nürnberg, pp. 1630-1633, 2015

[26] M. Frank and F. Zotter: „Exploring the perceptual sweet area in Ambisonics“, *142nd AES Convention*, Berlin, Deutschland, May 2017

[27] M. Frank and A. Sontacchi: „Case Study on Ambisonics for Multi-Venue and Multi-Target Concerts and Broadcasts“, submitted to *Journal of the Audio Engineering Society*, 2017

[28] M. Frank and F. Zotter: „Extension of the generalized tangent law for multiple loudspeakers“, *DAGA Tagungsbericht*, Kiel, pp. 1081-1084, 2017

[29] H. V. Fuchs: „Schallabsorber und Schalldämpfer“, *Springer-Verlag*, 3. Edition, Berlin/Heidelberg, 2010

[30] M. A. Gerzon: „Periphony: Width-Height Sound Reproduction“, *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2-10, 1973

[31] M. A. Gerzon: „The design of precisely coincident microphone arrays for stereo and surround sound“, *prepr. L-20 of 50th Audio Eng. Soc. Conv.*, 1975

[32] M. Gerzon: „Design of Ambisonics Decoders for for Multispeaker Surround Sound“, *58th Convention of the Audio Engineering Society*, New York, NY, Nov. 1977

[33] M. Gerzon: „General metatheory of auditory localisation“, *92nd Convention of the Audio Engineering Society*, Wien, März 1992

[34] M. Gräf and D. Potts: „On the computation of spherical designs by a new optimization approach based on fast spherical fourier transforms“, *Numer. Math.*, vol. 119, 2011, [Online] Available: http://homepage.univie.ac.at/manuel.graef/quadrature.php

[35] H. Haas: „Über den Einfluss eines Einfachechosauf die Hörsamkeit von Sprache“, *Acustica 1*, 1951, pp. 49-58

[36] R. H. Hardin and N. J. A. Sloane: „Mclaren's improved snub cube and other new spherical designs in three dimensions“, *Discrete and Computational Geomoetry*, vol. 15, pp. 429-441, 1996, [Online] Available: http://neilsloane.com/sphdesigns/dim3/

[37] J. M. Helm, E. Kurz and M. Frank: „Frequency-dependent amplitude-panning curves“, *29th Tonmeistertagung*, Cologne, Germany, November 2016

[38] Recommendation ITU-R BS.1116-1, „Methods for the subjective assessment of small impairments in audio systmes including multichannel sound systems“

[39] Recommendation ITU-R BS.2051-0, „Advanced sound system for programme production“, Genf, 2014

[40] L. A. Jeffress: „A place theory of sound localization“, *Journal of comparative and physiological psychology*, vol. 41, pp. 35-39, 1948

[41] M. Kronlachner: „Räumliche Transformationen zur Veränderung von Ambisonics Aufnahmen“, *Masterarbeit*, Universität für Musik und darstellende Kunst Graz, 2014

[42] M. Kleiner, J. Tichy: „Acoustics Of Small Rooms“, *Taylor and Francis Group*, Boca Raton, 2014

[43] E. Kurz: „Vorhersage des Hörbereichs für Surround-Wiedergabesysteme auf Basis des Energievektormodells“, *Toningenieur-Projekt*, Institut für Elektronische Musik und Akustik, 2017

[44] D. M. Leakey and E. Colin Cherry: „Influence of Noise upon the Equivalence of Intensity Differences and Small Time Delays in Two-Loudspeaker Systems“, *The Journal of the Acoustical Society of America*, vol. 29, no. 2, pp. 284-286, 1957

[45] D. M. Leakey: „Some measurements on the effects of interchannel intensity and time differences in two channel sound systems“, *The Journal of the Acoustical Society of America*, vol. 31, no. 7,pp. 977-986, 1959

[46] H. Lee and F. Rumsey: „Level and time panning of phantom images for musical sources“, *Journal of the Audio Engineering Society*, vol. 61, no. 12, pp. 978-988, 2013

[47] W. Lindemann: „Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals“, *The Journal of the Acoustical Society of America*, vol. 80, no. 6, pp. 1608-1622, 1986

[48] W. Lindemann: „Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals“, *The Journal of the Acoustical Society of America*, vol. 80, no. 6, pp. 1623-1630, 1986

[49] R. Y. Litovsky, H. S. Colburn, W. A. Yost and S. J. Guzman: „The Precedence Effect“, *The Journal of the Acoustical Society of America*, vol. 106, no. 4 Pt. 1, pp. 1633-1654, Oktober 1999

[50] J. P. A. Lochner and J. F. Burger: „The subjective masking of short time delayed echoes their primary sounds and their contribution to the intellegibility of speech“, *Acustica*, vol. 8, pp. 1-10, 1958

[51] D. G. Malham: „Experience with large area 3-D ambisonic sound system“, *Proceedings of the Institute of Acoustics*, vol. 14, no. 5, pp. 209-216, 1992

[52] G. Martin: „Microphone Techniques for Stereo and Multichannel“, *AES Tutorial 73*, 121st AES Convention, San Francisco, 2006

[53] E. Meyer and G. R. Schodder: „Über den Einfluß von Schallrückwürfen auf Richtungslokalisation und Lautstärke bei Sprache“, *Nachr. Akad. Wiss. in Göttingen*, Math. Phys. Klasse IIa, Vendenhoeck und Rupprecht, Göttingen, H. 6, pp. 31 - 42

[54] H. Mertens: „Directional hearing in stereophony“, *E.B.U. Review - Part A Technical*, no. 92, pp. 147 - 158, 1965

[55] C. Moldryzk, A. Goertz, M. Makarski, S. Feistel, W. Ahnert and S. Weinzierl: „Wellenfeldsynthese für einen großen Hörsaal“, *DAGA Tagungsbericht*, Stuttgart, 2007

[56] V. Pulkki: „Virtual sound source positioning using vector base amplitude panning“, *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456-466, 1997

[57] V. Pulkki and M. Karjalainen: „Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning“, *Journal of the Audio Engineering Society*, vol. 49, no. 9, pp. 739 - 752, September 2001

[58] M. S. Puckette: „Pure Data: another integrated computer music environment“, *Reprinted from Proceedings, Second Intercollege Computer Music Concerts*, pp. 37 - 41, Tachikawa, Japan, May 1997

[59] J. Ootomo, K. Tanno, A. Saji, J. Huang and W. Hatano: „The precedence effect of sound from a side direction“, *30th AES International Conference*, Saariselkä, Finland, March 2007

[60] B. Rakerd and W. M. Hartmann: „Localization of sound in rooms, II: The effects of a single reflecting surface“, *Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149 - 2157, 2004

[61] B. Rakerd, W. M. Hartmann and J. Hsu: „Echo suppression in the horizontal and median sagittal planes“, *The Journal of the Acoustical Society of America*, vol. 107, pp. 1061 - 1064, 2000

[62] P. W. Robinson, A. Walther, C. Faller and J. Braasch: „Echo thresholds for reflections from acoustically diffusive architectural surfaces“, *The Journal of the Acoustical Society of America*, Vol. 134, pp. 2755 - 2764, 2013

[63] W. C. Sabine: „Collected papers on acoustics“, *Harvard University Press*, Cambridge, 1922

[64] E. Sengpiel and tonmeister students: „Direction of Phantom sources - Stereo loudspeaker signals - 1 Interchannel level differences and interchannel time differences Localization Curves“, Berlin, 1992, [Online] Available: http://www.sengpielaudio.com/InterchannelLevelDifferencesAndInterchannelTimeDifferences1.pdf

[65] H. P. Seraphim: „Über die Wahrnehmbarkeit mehrerer Rückwürfe von Sprachschall“, *Acustica 11*, pp. 80 - 91

[66] B. G. Shinn-Cunningham, P. M. Zurek and N. I. Durlach: „Adjustment and discrimination measurements of the precedence effect“, *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2923-2932, 1993

[67] G. Simonsen: „Master's Thesis“, *Technical University of Lyngby*, Denmark, 1984

[68] P. Stitt: „Ambisonics and Higher-Order Ambisonics for Off-Centre Listeners: Evaluation of Perceived and Predicted Image Direction“, *Doctoral Thesis*, Queen's University Belfast, April 2015

[69] P. Stitt, S. Bertet and M. van Walstijn: „Off-Center Listening with Third-Order Ambisonics: Dependence of Perceived Source Direction on Signal Type“, *Journal of the Audio Engineering Society*, vol. 65, no. 3, March 2017

[70] S. Tervo, J. Pätynen, A. Kuusinen and T. Lokki: „Spatial decomposition method for room impulse responses“, *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17 - 28, 2013

[71] H. Wallach, E. B. Newman and M. R. Rosenzweig: „The precedence effect in sound localization“, *The American Journal of Psychology*, vol. 62, no. 3, pp. 315-336, April 1949

[72] D. Ward and T. Abhayapala: „Reproduction of a plane-wave sound field using an array of loudspeakers“, *IEEE Transactions on Speech and Audio Processing*, 9(6), pp. 697-707, 2001

[73] H. Warncke: „Die Grundlagen der raumbezüglichen stereophonischen Übertragung im Tonfilm“, *Akust. Z. 6*, pp. 174 - 188, 1941

[74] K. Wendt: „Versuche zur Ortung von Intensitäts-Stereophonie“, *FREQUENZ*, no. 1, pp. 11 - 14, 1960

[75] K. Wendt: „Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie“, *Ph.D. dissertation*, RWTH Aachen, 1963

[76] F. Wendt, M. Frank and F. Zotter: „Application of localization models for vertical phantom sources“, *AIA-DAGA Konferenzband-Beitrag*, Berlin, 2013

[77] F. Wendt, M. Frank and F. Zotter: „On the localization of auditory objects created by directional sound sources in a virtual room“, *30th Tonmeistertagung*, Cologne, Germany, November 2018

[78] J. Zmölnig: „Entwurf und Implementierung einer Mehrkanal-Beschallungsanlage“, *Diplomarbeit*, Universität für Musik und darstellende Kunst Graz, 2002

[79] F. Zotter: „Analysis and Synthesis of Sound-Radiation with Spherical Arrays“, *Doktorarbeit*, Institut für Elektronische Musik und Akustik, Graz, 2009

[80] F. Zotter: „Sampling strategies for acoustic holography/holophony on the sphere“, in *NAG-DAGA*, Rotterdam, 2009

[81] F. Zotter, M. Frank, A. Fuchs and D. Rudrich: „Preliminary study on the perception of orientation-changing directional sound sources in rooms“, *Proceedings of Forum Acusticum*, Krakau, September 2014

[82] F. Zotter and A. Sontacchi: „Icosahedral Loudspeaker Array“, *IEM Report 39/07*, Institut für Elektronische Musik und Akustik, Graz 2007

[83] F. Zotter and M. Frank: „All-Round Ambisonic Panning and Decoding“, *Journal of the Audio Engineering Society*, vol. 60, no. 10, October 2012

[84] F. Zotter and M. Frank: „Generalized tangent law for horizontal pairwise amplitude panning“, *Proceedings from International Conference On Spatial Audio*, Graz, September 2015

[85] F. Zotter and M. Frank: „Ambisonic decoding with panning-invariant loudness on small layouts (AllRAD2)“, *144th AES Convention*, Milan, Italy, 2018