Phantom Sources using Multiple Loudspeakers in the Horizontal Plane

Dissertation by

Matthias Frank

Assessors:
Professor Robert Höldrich (IEM, KUG)
Professor Jens Blauert (IKA, RUB)

Institute of Electronic Music and Acoustics
University of Music and Performing Arts Graz, Austria
June 2013

Matthias Frank (Name in Blockbuchstaben)

0430270 (Matrikelnummer)

Erklärung

Hiermit bestätige ich, dass mir der *Leitfaden für schriftliche Arbeiten an der KUG* bekannt ist und ich diese Richtlinien eingehalten habe.

Graz, den …….

……
Unterschrift der Verfasserin / des Verfassers

ABSTRACT

This work investigates phantom sources created by amplitude panning in the horizontal plane. Although there exist models about the perception of phantom sources in literature, their experimental evidence for panning methods that use more than two simultaneously active loudspeakers is rare. The listening experiments in this work contribute to close this gap by investigating localization, source width and coloration of phantom sources using multiple loudspeakers in the horizontal plane. All experiments are performed using pink noise in a typical studio environment. Besides the comparison of the experimental results to existing models, some models are adapted to improve their applicability, and new models are developed.

KURZFASSUNG

Diese Arbeit untersucht Phantomschallquellen, die bei Amplitudenpanning in der Horizontalebene entstehen. Obwohl in der Literatur Modelle für die Wahrnehmung von Phantomschallquellen existieren, ist ihre experimentelle Bestätigung für die Verwendung von mehr als zwei gleichzeitig aktiven Lautsprechern rar. Die Hörversuche in dieser Arbeit tragen zur Schließung dieser Lücke bei, indem sie Lokalisation, Breite und Klangfarbe von Phantomschallquellen unter Verwendung von mehreren Lautsprechern in der Horizontalebene untersuchen. Alle Versuche werden mit rosa Rauschen unter typischen Studioabhörbedingungen durchgeführt. Neben dem Vergleich der Versuchsergebnisse zu bestehenden Modellen werden einige Modelle angepasst, um ihre Eignung zu verbessern, sowie neue Modelle entwickelt.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors, Robert Höldrich, Jens Blauert, and Gerhard Eckel for fruitful discussions that always energized my work on this thesis.
I am very grateful for the employment at the Institute of Electronic Music and Acoustics (IEM), University of Music and Performing Arts Graz (KUG) during the last four years and the possibility of using the extensive facility and equipment at IEM. In this respect, my thanks go the Alois Sontacchi, the head of IEM.
I thank the whole IEM staff for the friendly working environment and interesting discussions. Special thanks go to my colleagues of the former office in the 1st floor for the great atmosphere and the daily lunch.
I would particularly like to thank Franz Zotter, who spent evenings and weekends on discussions, (co-)authoring of papers, and proof-reading of my articles and this thesis.
I also thank the students that prepared the way for my thesis by preliminary experiments in their seminar or student works.
The experiments would not have been possible without participants. I therefore thank all members of the expert listening panel for their participation and the AAP [1] project, as well as the IEM for absorbing the expense allowance.

For the appreciated distraction and diversion from the scientific work and the encouragement of creativity, I thank all the musical formations I take part in, especially the Grazer Bläservielharmonie, the staff choir of the KUG, and my band einheits_brei.

Finally, I would like to express my gratitude to my parents Peter and Doris, as well as my brother Andreas for their ongoing support and motivation.

[1] The project AAP is funded by Austrian ministries BMVIT, BMWFJ, the Styrian Business Promotion Agency (SFG), and the departments 3 and 14 of the Styrian Government. The Austrian Research Promotion Agency (FFG) conducted the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a programme of the above-mentioned institutions.

Contents

Chapter I

INTRODUCTION

In a typical listening environment, sound playback on a single loudspeaker causes a single auditory event that is perceived at the location of the loudspeaker, cf. left side of Figure 1. *Stereophony* [Blu58] plays back the same sound over a pair of loudspeakers and creates an auditory event that can be perceived between the loudspeakers, cf. center of Figure 1. This auditory event is called *phantom source* [Wen63, The80]. Its position can be controlled by the relation of the loudspeaker gains g_1 and g_2, often referred to as *amplitude panning*. However, does this phantom source have the same quality as an auditory event caused by a single loudspeaker? Does it have the same timbre and the same apparent width?

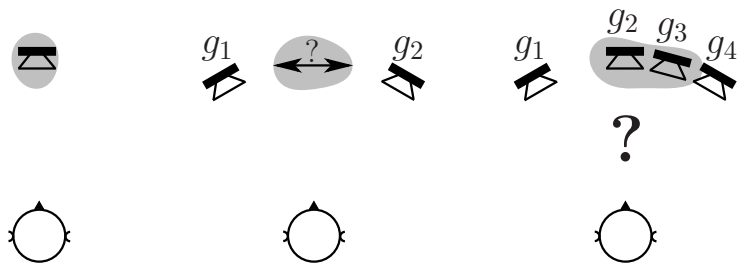


Figure 1: Auditory events (gray) caused by different numbers of active loudspeakers and gains g_l.

What happens if more loudspeakers are used for the creation of a single phantom source, cf. right side of Figure 1? Does this even work? How can all the gains $g_1, \ldots g_4$ be used to control the position of a single phantom source?
This work investigates these questions by studying phantom sources that are created by amplitude panning on multiple loudspeakers in the horizontal plane. It does not only study the control of the phantom source location, but also its side effects, i.e. changes in the width and coloration of the phantom source. For the studies, own listening experiments are conducted under typical non-anechoic studio conditions using pink noise as stimulus. Results from these experiments and the literature are compared to existing models for phantom source perception. Some of these models are adapted and new models are developed.

Other Spatialization Methods

Besides level differences, time-delay differences can be used for the panning of a phantom source. However, amplitude panning is most common as the localization

is more accurate [Lea59, Wen63, Bla83, Pul01b].
Other spatialization methods do not aim at creating an auditory event that is similar to the one caused by a real sound source, but they reconstruct the physical properties of the sound field of a real source. They often use the concept of a *virtual source*, i.e. a computational representation for properties of the real sound source.
Binaural methods recreate the ear signals that would result from a real source at the desired location using headphones. Although this technique creates the natural binaural cues of humans that are essential for a proper spatial impression [Bla83], it requires a fast dynamic adaptation to the listener's head orientation [Mac08, LPLP12] in order to create convincing results. For one head orientation and one virtual source, a convolution of the source signal with two so-called head-related impulse responses is necessary. The distance of the auditory event can be controlled by the selection of appropriate impulse responses. These methods can also be applied to loudspeakers [CB89] using crosstalk cancellation [PR11]. As binaural methods do not require much hardware effort, they are used to simulate other reproduction methods that require a large number of loudspeakers [Völ11, WS12].
Holophony or *Sound Field Synthesis* aim at recreating the extended entire sound field of a virtual source by utilizing a large number of loudspeakers. Basically, various types of virtual sources can be synthesized, such as plane waves or point sources at arbitrary positions. For one virtual source, the source signal has to be convolved with a specific impulse response for each loudspeaker. Although the theoretical basis was known before [SS34, Sno53, Jes73], the first practical systems with a large number of loudspeakers and suitable computational power were built in the late eighties [Ber88]. The two best-known sound field synthesis methods are *Wave Field Synthesis* [BdVV93, SRA08] and *Near-Field Compensated Higher-Order Ambisonics* [Dan03, FNCS08, WA09]. Typical systems consist of several hundred loudspeakers. However, none of the existing systems provides loudspeaker spacings that can yield a physically accurate synthesis up to 20kHz. The perceptual implications of larger loudspeaker spacings are currently under investigation [Wit07, WRS12].

In order to clearly distinguish between the different methods of creating auditory events, the term phantom source in this works exclusively refers to auditory events that are created by amplitude panning.

1.1 Modeling of Phantom Sources

The localization of a single sound source in the horizontal plane is primarily based on binaural differences, i.e. differences between the ear signals. These differences are composed of inter-aural level differences (ILDs) and inter-aural time differences (ITDs) and can be measured by the head-related transfer functions (HRTFs) [Bla83]. The ILDs are mainly responsible for localization of high frequencies, the ITDs for lower frequencies [Str07]. Newer studies revealed that for broadband sources, low frequency ITDs are dominant [WK92] and that ITDs are also important for higher frequencies [MM02].

Binaural localization models evaluate ITD and ILD values and compare them to a database in order predict the direction of a sound source, see also Section 3.2.1. However, the models can hereby only predict lateralization, i.e., how far the auditory event is perceived towards left or right. For the discrimination between front and back, monaural spectral cues are crucial [LB02]. The spectral cues arise from the frequency-dependent directivity of human hearing, e.g. the attenuation of high frequencies for dorsal directions, see also Figure 17(a). The perception can utilize these cues only by separating or estimating the respective spectral cue within the auditory event, presumably by detecting or having memory of a reference spectrum underlying the auditory object.

All above-mentioned mechanisms work for single real sound sources. This section introduces different models about the perception of phantom sources.

1.1.1 Summing Localization

The upper part of Figure 2 shows the principle of summing localization [Bla83]. The sound from multiple loudspeakers linearly superimpose at the ears of the listener. Summing localization assumes that the characteristics of the superimposed sound field are similar to the characteristics of the sound field that is produced by a single real source. A listener perceives a single auditory event at the location of this virtually equivalent single sound source, a phantom source.

The above-mentioned characteristics are the binaural localization cues, i.e. ITD and ILD. Level differences between the loudspeakers at low frequencies create ITDs [Lea59, Wen63, Bla83, Pul01b]. This is illustrated in Figure 3 for a pair of loudspeakers at $\pm 30°$. As the ILD can be neglected for low frequencies, the sound of each loudspeaker reaches both ears with the same magnitude $|p_{1,left}| = |p_{1,right}|$ and $|p_{2,left}| = |p_{2,right}|$. Adjusting the level of the loudspeakers to $g_1 = 1$ and $g_2 = 0.25$ results in the same magnitude of the loudspeaker signals at both ears. However, there is an ITD due to the positions of the loudspeakers. The ITD linearly converts to a phase difference (IPD) at low frequencies. It can be read

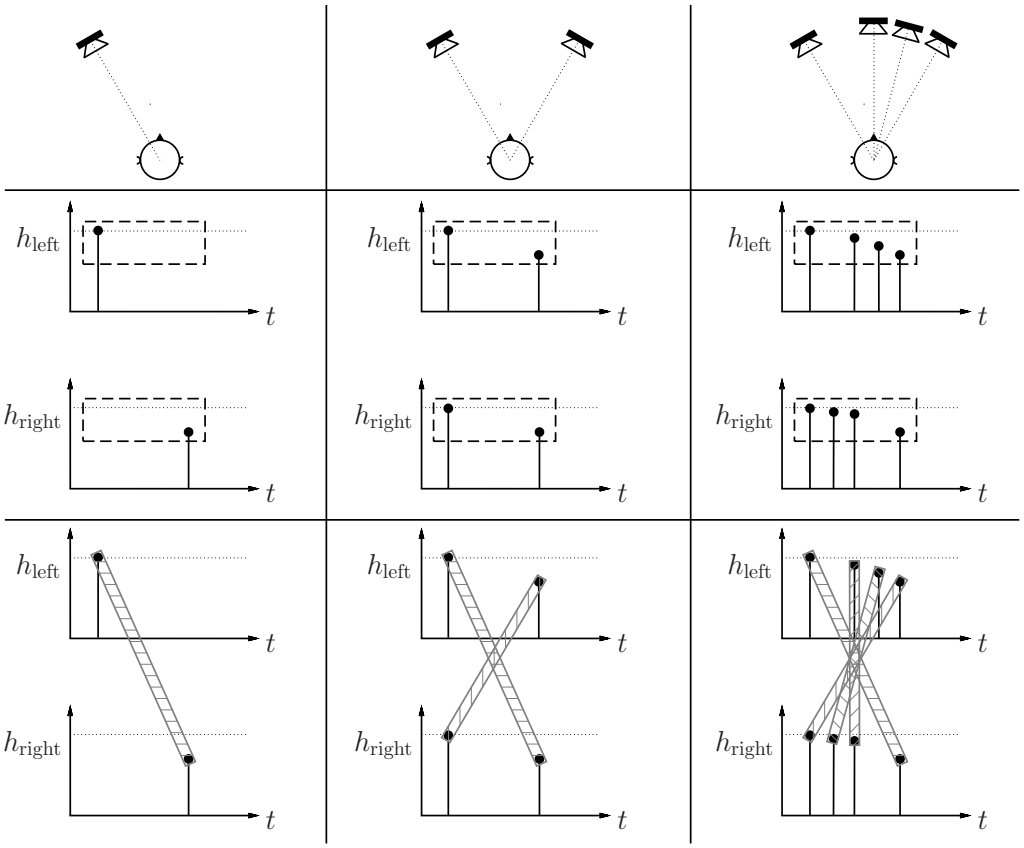


Figure 2: Simplified impulse responses at the ears for 1, 2, and 4 loudspeakers. Impulses are summarized according to summing localization (black dashed rectangles) and association model (gray hatched rectangles). Figure adapted from [The80].

off the sound pressure phasors sketched in Figure 3 for a symmetric loudspeaker arrangement. Summing up the phasors at each ear yields p_{left} and p_{right} with $|p_{left}| = |p_{right}|$ (zero ILD) and $\angle(p_{left}) \neq \angle(p_{right})$ (nonzero IPD or ITD). As low-frequency ITDs are the dominant localization cues [WK92], amplitude panning delivers quite accurate localization at the central listening position.

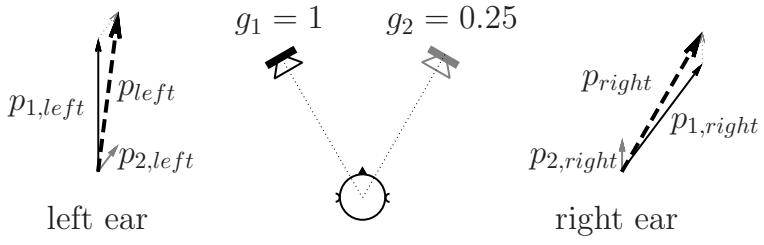


Figure 3: Creation of ITDs by level differences at low frequencies. Figure adapted from [Bla83].

Summing localization does not assume a certain number of simultaneously active sound sources [Bla83]. It is the basis of the computational binaural localization models that is used in this thesis. Moreover, the binaural model of coloration and

the source width measure using the inter-aural coherence, applied in this work, are also based on the idea of summing localization.

1.1.2 Association Model

The association model by Theile [The80] does not only deal with localization, it also covers other properties of phantom sources, such as coloration. It assumes two different processing mechanisms for the auditory system that are both based on pattern recognition and association. The first stage yields a location association and the second stage a gestalt association. The stages deliver two independent attributes of the auditory event, namely location and gestalt, which always appear as an output pair.

The basic difference to summing localization is the assumption that the loudspeakers contributing to a phantom source can be localized separately by the location association stage, cf. lower part of Figure 2. This stage applies an inverse filtering with the HRTFs corresponding to the localized loudspeaker positions. The filtered output is fed to the gestalt association stage, where the loudspeaker signals are compared and fused to a single phantom source. Due to the inverse HRTF filtering, the association model provides explanations for binaural decoloration [Sal95, Brü01], i.e., the phenomenon that coloration is suppressed when listening with both ears. In [The80] it is stated that this also works for phantom sources although physically, the superposition of the loudspeaker signals creates a comb filter at each ear.

All examples in the work about the association model [The80] use a maximum of two simultaneously active loudspeakers. It is not known how many loudspeakers can be separated by the localization association stage. Up to now, there exists no technical implementation of the association model. This is most probably due to the necessary association and inverse HRTF filtering that are impractical for a technical implementation of finite complexity.

1.1.3 Vector Models

Another option of modeling phantom sources are vector models. These models are not based on binaural signals but on properties of the sound field at the listening position. The velocity vector can be seen as a normalized plane-wave decomposition of 1^{st} order (monopole, dipoles) [Mak62, Ger92]. It is assumed to predict localization for low frequencies and can be computed solely based on loudspeaker positions and gains. The 1^{st} order components can also be measured in a real sound field using a coincident arrangement of a pressure microphone (omni-directional) and pressure difference microphones (figure-of-eight) aligned

with the axis of the vector space. The intensity vector can be computed from these measurements and indicates the direction of the sound field at the measurement position [Hey86]. Both the velocity vector and the intensity vector include the phase relation between the components.
Following the properties of human ears that phase relations cannot be dissolved at higher frequencies, the energy vector describes the direction and distribution of energy and is considered as a model of localization at higher frequency [Ger92]. Similar to that, the lateral energy fraction measure quantifies the amount of energy from lateral directions in relation to the energy from all directions and is known as a measure for the apparent source width in room acoustics.
Although the suitability of vector measures for the prediction of phantom source perception has not been evaluated yet, they are often applied in practice [MW08, MW10]. In this thesis, their suitability is investigated and compared to the state-of-the-art predictors that are based on the analysis of binaural signals.

1.2 Organization of Contents

The second chapter introduces the amplitude-panning methods that are part of the evaluation in this work, namely Vector-Base Amplitude Panning (VBAP), Multiple-Direction Amplitude Panning (MDAP), and Ambisonics. All experiments in this work use the same experimental setup. This setup is also described in the second chapter.
The third chapter studies the main goal of panning, i.e. the control of the phantom source location. As the numerical details of different panning methods may not be a good starting point for understanding localization mechanisms in multiple-loudspeaker panning, the study starts with a literature review of experimental results about pairwise panning in the standard $\pm 30^\circ$ loudspeaker setup. The study also considers pairwise panning for arbitrary loudspeaker positions in the horizontal plane. Own experiments extend the localization study to multiple loudspeakers using VBAP, MDAP, and Ambisonics. In addition to the comparison of prediction models for localization, the chapter presents experimental results for an off-center listening position. The chapter briefly discusses three-dimensional panning and gives a possible explanation for subjective variation in phantom source localization.
Phantom source width is studied in the fourth chapter. The study starts with a generic experiment using one, two, and three simultaneously active loudspeakers in a symmetric arrangement. The experimental results are compared to source width measures from room acoustics. On the one hand, a modification of these measures

for phantom source width is introduced and on the other hand a simple predictor. The chapter subsequently extends the generic study to amplitude-panning methods using multiple loudspeakers, such as VBAP, MDAP, and Ambisonics, by further experiments. It briefly presents an outlook on phantom source width in three-dimensional panning and phantom source widening by decorrelation.
The fifth chapter discusses phantom source coloration. Although coloration is typically evaluated by verbal attributes, this section uses a spectral binaural model for the quantitative description of coloration changes. This model is used to study the generic cases of two and three simultaneously active loudspeakers, as well as typical amplitude-panning scenarios. The model predictions are compared to results of a listening experiment that was done to investigate coloration of moving phantom sources using different amplitude-panning methods and listening positions. Finally, the chapter proposes simpler predictors of coloration changes in amplitude panning that are solely based on the loudspeaker positions and gains.
The sixth and last chapter concludes the thesis by summarizing and comparing the results of the chapters.

Parts of this work have already been published. This applies to the pointing method in Section 2.3 [FMSZ10], the localization experiments in Section 3.3.3 [BF11, WFZ13], the study about source width in Section 4.1 [Fra13], experiment 2 in Section 4.2 [FMS11], and the study about phantom source widening using decorrelation [ZFMS11, ZF13, FZ13]. The content of Sections 3.1 and 4.2 has been submitted to a journal.

Chapter II

PANNING METHODS AND EXPERIMENTAL SETUP

This chapter gives a short overview about the amplitude-panning methods under investigation. Moreover, it presents the experimental setup and the pointing method used for the localization experiments.
In this work, the directions of L loudspeakers and the panning directions are expressed as vectors of unit length $\boldsymbol{\theta} = [\cos(\phi),\ \sin(\phi)]^{\mathrm{T}}$ that depend on the azimuth angle ϕ in the horizontal plane, see Figure 4. For each loudspeaker $l \in \{1 \ldots \mathrm{L}\}$, the scalar weight g_l denotes its adjustable gain. Impulse responses are represented by $h(t)$. Non-italic symbols express constant values or labels.

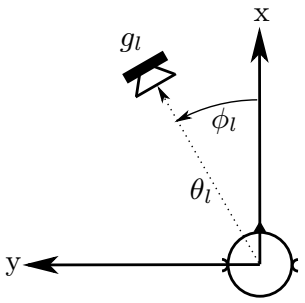


Figure 4: Sketch of the direction variables in the reference system, taking the lth loudspeaker as an example.

2.1 Amplitude-Panning Methods

In contrast to sound field synthesis, amplitude panning can only render point sources on the radius of the loudspeaker arrangement. Nevertheless, a distance impression can be created by perceptual cues, such as a change in level and the direct-to-reverberation ratio [Zah02]. Although all presented amplitude-panning methods are also applicable to three-dimensional loudspeaker arrangements, the overview focuses on the two-dimensional case as it is evaluated in the remainder of this work.

2.1.1 Vector-Base Amplitude Panning (VBAP)

Vector-Base Amplitude Panning (VBAP) [Pul97] is the generalization of the *tangent law* [Lea59] for panning of phantom sources [Wen63] in two-channel stereophony [Blu58]. The tangent law is based on a simple geometrical head model and is the most popular panning law for pairwise panning, cf. Section 3.1.1. In the horizontal case, VBAP calculates two weights $\boldsymbol{g}_{\mathrm{ij}} = [g_{\mathrm{i}},\, g_{\mathrm{j}}]^{\mathrm{T}}$ applicable to creating the impression of a phantom source between two loudspeakers that are located at $\mathbf{L}_{\mathrm{ij}} = [\boldsymbol{\theta}_{\mathrm{i}},\, \boldsymbol{\theta}_{\mathrm{j}}] = \begin{bmatrix} \cos(\phi_{\mathrm{i}}) & \cos(\phi_{\mathrm{j}}) \\ \sin(\phi_{\mathrm{i}}) & \sin(\phi_{\mathrm{j}}) \end{bmatrix}$.
The unnormalized weights $\tilde{\boldsymbol{g}}_{\mathrm{ij}}$ are calculated from the panning direction $\boldsymbol{\theta}_s$ by

$$\mathbf{L}_{\mathrm{ij}}\, \tilde{\boldsymbol{g}}_{\mathrm{ij}} = \boldsymbol{\theta}_s \qquad\qquad \Rightarrow \tilde{\boldsymbol{g}}_{\mathrm{ij}} = \mathbf{L}_{\mathrm{ij}}^{-1}\, \boldsymbol{\theta}_s, \tag{1}$$

and subsequent normalization yields the weights $\boldsymbol{g}_{\mathrm{ij}} = \tilde{\boldsymbol{g}}_{\mathrm{ij}} / \|\tilde{\boldsymbol{g}}_{\mathrm{ij}}\|$ for panning. In order to give good results, the aperture of the loudspeaker pair must not exceed 90° [Pul01b], i.e., for source directions lying inside the loudspeaker pair expressed by $\mathbf{L}_{\mathrm{ij}}$, the weights must be positive. To extend the panning range around this pair, more loudspeaker pairs need to be attached [Pul97].

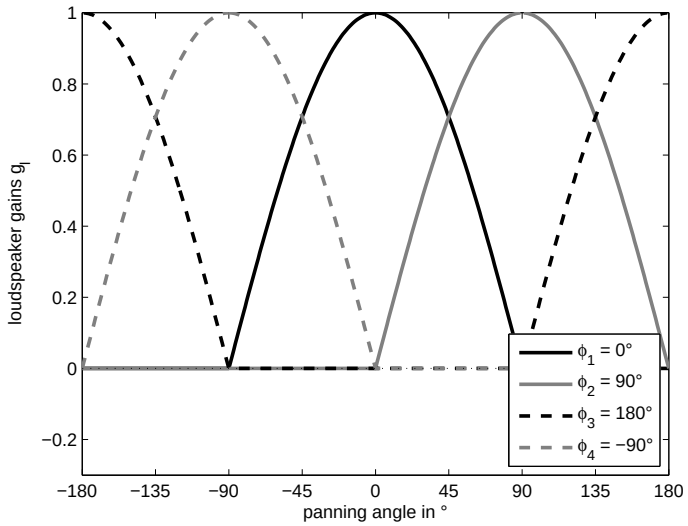


Figure 5: Gains for different panning angles using VBAP on a regular arrangement of 4 loudspeakers with $\Delta\phi_L = 90°$.

Figure 5 shows the panning gains for the 4 loudspeakers of a regular arrangement using VBAP. The number of active loudspeakers is depending on the panning direction: two loudspeakers are active for directions between two loudspeakers and one is active for directions coinciding with a loudspeaker.

2.1.2 Multiple-Direction Amplitude Panning (MDAP)

For a more uniform panning, those cases of VBAP with only one active loudspeaker should be avoided. This can be done by extending VBAP to Multiple-Direction Amplitude Panning (MDAP) [Pul99].
For a desired panning angle ϕ_s, MDAP superimposes the results of VBAP for B panning directions uniformly distributed around the desired panning direction within a spread of $\pm\phi_{\mathrm{MDAP}}$. Typically, the maximum spread is related to the loudspeaker spacing $\Delta\phi_L$ of a uniform loudspeaker arrangement: $\phi_{\mathrm{MDAP}} = 1/2\Delta\phi_L$. For this setting, three loudspeakers are active for most panning directions, even for directions on a loudspeaker. If the panning direction lies exactly in the middle between two loudspeakers, only two loudspeakers are active. Only in these special cases, MDAP yields the same results as VBAP.
In this work, MDAP is used with B = 10 panning directions uniformly distributed within a spread of $\phi_{\mathrm{MDAP}} = 1/2\Delta\phi_L$.

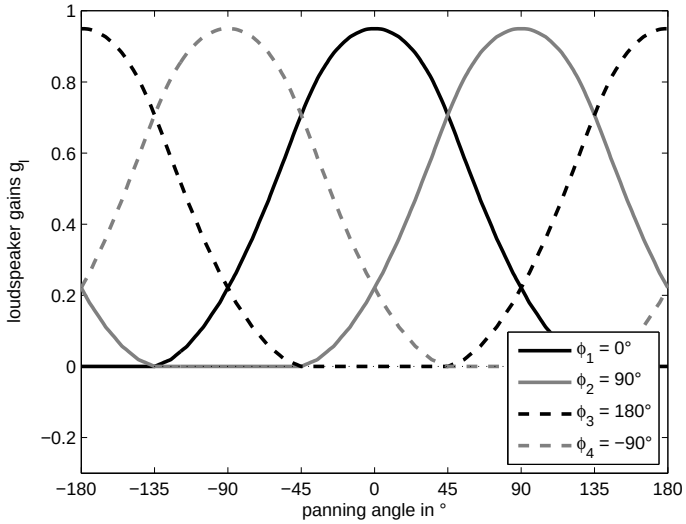


Figure 6: Gains for different panning angles using MDAP on a regular arrangement of 4 loudspeakers with $\Delta\phi_L = 90°$, B=10, and $\phi_{\mathrm{MDAP}} = 45°$.

Figure 6 shows the loudspeaker gains for MDAP on a regular array of 4 loudspeakers. In the example, the 10 panning directions for a desired direction of $\phi_s = 0°$ lie at $-45°, -35°, \ldots, 45°$. The number of active loudspeakers is increased in comparison to VBAP.

2.1.3 Ambisonics

Ambisonics [CS72, Ger73, Dan01] is a recording and reproduction method which is based on the representation of the sound field excitation as a superposition of orthogonal basis functions. In the horizontal case, they correspond to the periodic trigonometric basis of the Fourier series, the *circular harmonics*. Their maximum order N determines the spatial resolution and the number $2\mathrm{N}+1$ of signals and minimum required loudspeakers.
Although Ambisonics is also used for sound field synthesis [Dan03, AS08, FNCS08, WA09, Fra09], this work applies it as a mere amplitude-panning method [CS72, Ger75, Fel75, MM95, Pol96, Cra03].

Encoding. For one source at a direction $\boldsymbol{\theta}_s = [\cos(\phi_s),\ \sin(\phi_s)]^{\mathrm{T}}$, the *encoder* computes the Ambisonic spectrum $\boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_s)$ by evaluating the circular harmonics at $\boldsymbol{\theta}_s$

$$\boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_s) = [1/\sqrt{2},\ \cos(\phi_s),\ \sin(\phi_s),\ \cos(2\phi_s),\ \sin(2\phi_s),\ \ldots\ ,\ \cos(\mathrm{N}\phi_s),\ \sin(\mathrm{N}\phi_s)]^{\mathrm{T}}\,. \tag{2}$$

The assumption that all sources in the sound field and the reproduction system lie on a circle of the same radius r yields a frequency-independent processing.

Decoding. The *decoder* derives the gains $\boldsymbol{g} = \{g_1, ...g_{\mathrm{L}}\}$ for the L loudspeakers of an arrangement from the Ambisonic spectrum $\boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_s)$ by multiplying with the decoder matrix $\mathbf{D}$:

$$\boldsymbol{g} = \mathbf{D}\,\mathrm{diag}\{\boldsymbol{a}_{\mathrm{N}}\}\,\boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_s). \tag{3}$$

In order to control the main and side lobes emerging from the truncation of the circular harmonics, a weighting vector $\boldsymbol{a}_{\mathrm{N}}$ is applied in the harmonics domain [Dan01]. The *basic* or *max* $\boldsymbol{r}_{\mathrm{V}}$ weighting[1] uses a vector of ones $\boldsymbol{a}_{\mathrm{N}} = \mathbf{1}$, whereas the *max* $\boldsymbol{r}_{\mathrm{E}}$ weighting suppresses the side lobes at the cost of a wider main lobe by attenuating higher orders, cf. Table 1. Another weighting, called *in-phase*, yielded no convincing results in experiments [FZ08] and is therefore not used here. In the remainder of this work, basic Ambisonics is abbreviated with $\mathrm{Ambi}^{\mathrm{rV}}$ and Ambisonics using $\max \boldsymbol{r}_{\mathrm{E}}$ weighting with $\mathrm{Ambi}^{\mathrm{rE}}$.

Any decoder is derived from the circular harmonic spectra $\boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_l)$ of each loudspeaker

$$\boldsymbol{Y}_{\mathrm{N}} = [\boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_1),\ \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_2),\ \ldots,\ \boldsymbol{y}_{\mathrm{N}}(\boldsymbol{\theta}_{\mathrm{L}})]. \tag{4}$$

[1] Strictly speaking, $|\boldsymbol{r}_{\mathrm{V}}|$ is not maximized but constantly 1.

	$\max \boldsymbol{r}_\mathrm{V}$	$\max \boldsymbol{r}_\mathrm{E}$
weight $a(n)$	1	$\cos\left(\frac{n\pi}{2\mathrm{N}+2}\right)$

Table 1: Ambisonic weights: $\boldsymbol{a}_\mathrm{N} = [a(0),\, a(1),\, a(1),\, \ldots,\, a(\mathrm{N})]^\mathrm{T}$.

The decoder matrix $\mathbf{D}$ can be calculated by transposition or inversion of $\boldsymbol{Y}_\mathrm{N}$, resulting in a *sampling decoder* or *mode-matching decoder* [Pol00], respectively. The *energy-preserving decoder* [ZPN12] uses more sophisticated techniques, such as singular value decomposition. An overview about different decoders can be found in the appendix of [ZF12]. The computation of a suitable decoder matrix is a challenging task, especially if the loudspeaker arrangement is irregular or covers only a part of a circle. Decoders for circular arrangements on the $\mathrm{L} \geq 2\mathrm{N}_\mathrm{max} + 1$ vertices of regular polygons are sampling, mode-matching, and energy-preserving for all orders $\mathrm{N} \leq \mathrm{N}_\mathrm{max}$ due to the orthogonality of the uniformly discretized Fourier series. Therefore, this study uses such arrangements exclusively to reinforce the clarity of the undertaken experiments.

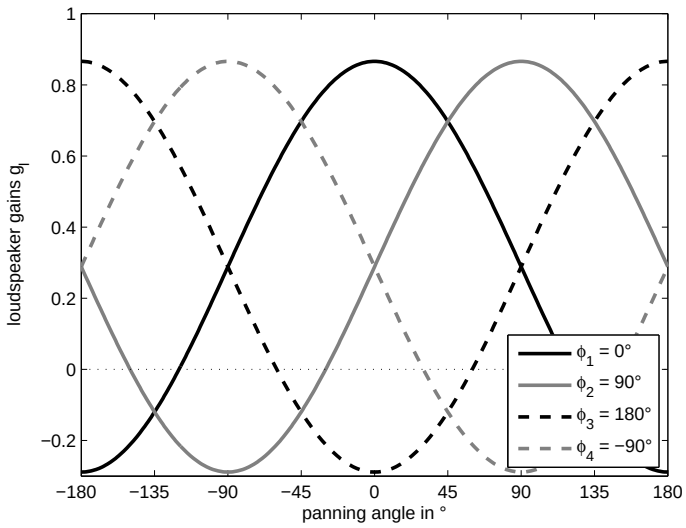


Figure 7: Gains for different panning angles using 1st order Ambi$^\mathrm{rV}$ on a regular arrangement of 4 loudspeakers.

Figure 7 and 8 show the loudspeaker gains using 1st order Ambi$^\mathrm{rV}$ and Ambi$^\mathrm{rE}$, respectively. Ambi$^\mathrm{rE}$ obviously attenuates the side lobes in comparison to Ambi$^\mathrm{rV}$. For Ambisonics, there exist suitable microphone arrays for recording [CG74, ME02].

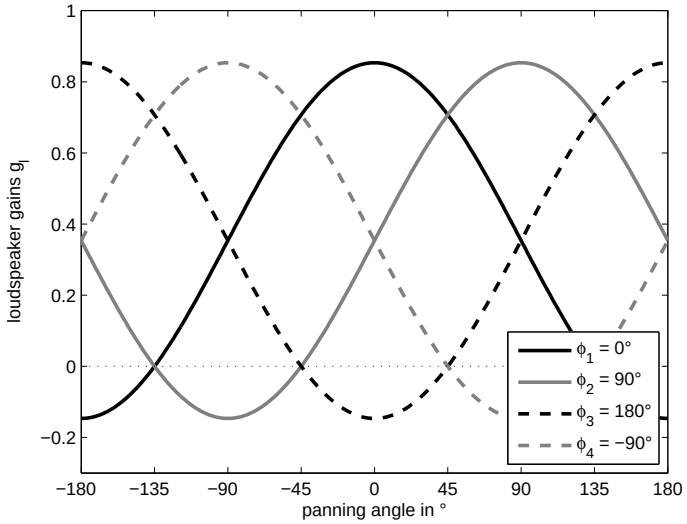


Figure 8: Gains for different panning angles using 1st order Ambi$^{\mathrm{rE}}$ on a regular arrangement of 4 loudspeakers.

2.1.4 Comparison

Figures 9, 10, and 11 compare the loudspeaker gains for the different panning methods. The gains are exemplarily shown for the first loudspeaker at $\phi_1 = 0°$.

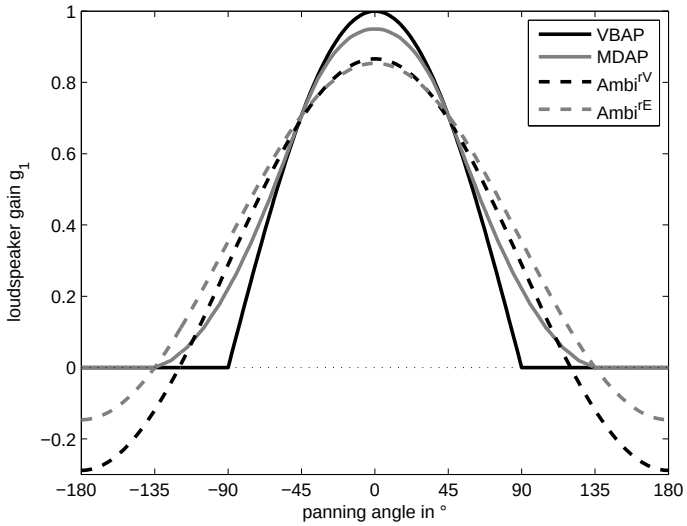


Figure 9: Comparison of panning methods using 4 loudspeakers (and 1st order Ambisonics) with $\Delta\phi_L = 90°$: gain of loudspeaker 1 at $\phi_1 = 0°$.

The overall characteristics of each panning method is independent of the number of loudspeakers and, in case of Ambisonics, of the order. VBAP always uses one or two loudspeakers and yields the smallest main lobe with no side lobes. In comparison the main lobe of MDAP is wider without creating side lobes. Side lobes are present when using Ambi$^{\mathrm{rV}}$ or Ambi$^{\mathrm{rE}}$. Obviously, the first side lobe

is the strongest one. Its level is not depending on the order and the number of loudspeakers for Ambi$^{\mathrm{rV}}$. However, for higher orders the first side lobe is closer to the main lobe. This results in an increased attenuation of sound from the direction that is opposite to the desired panning direction. The attenuation of the side lobes increase with increasing number of loudspeakers and order of Ambi$^{\mathrm{rE}}$.

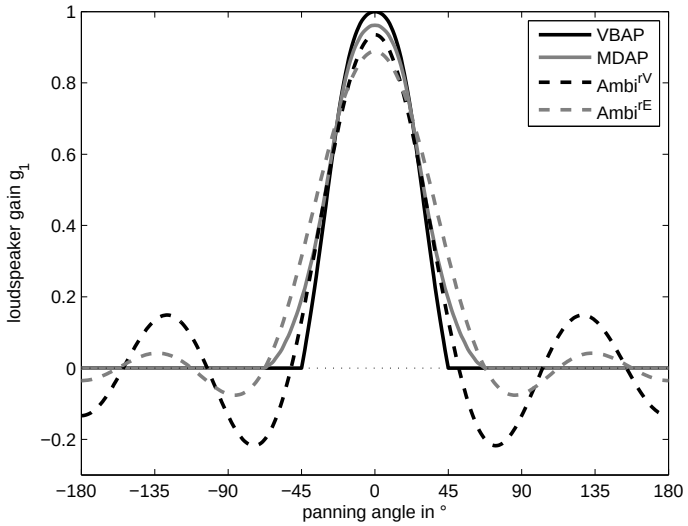


Figure 10: Comparison of panning methods using 8 loudspeakers (and 3$^{\mathrm{rd}}$ order Ambisonics) with $\Delta\phi_L = 45°$: gain of loudspeaker 1 at $\phi_1 = 0°$.

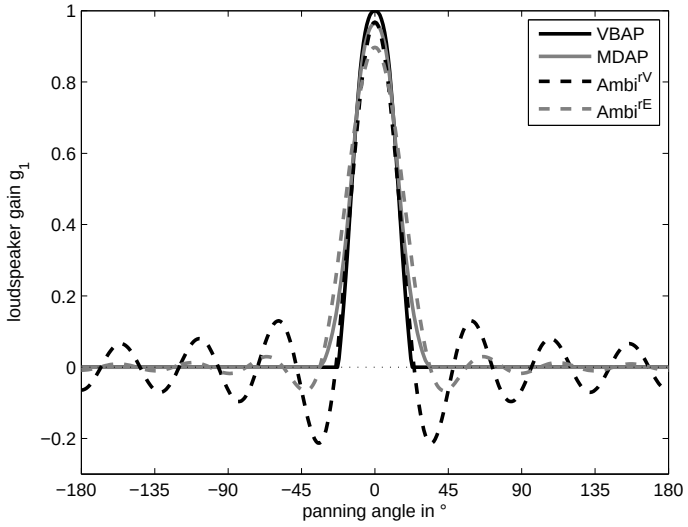


Figure 11: Comparison of panning methods using 16 loudspeakers (and 7$^{\mathrm{th}}$ order Ambisonics) with $\Delta\phi_L = 22.5°$: gain of loudspeaker 1 at $\phi_1 = 0°$.

2.2 Experimental Setup

All experiments performed in this work use the same experimental setup, except for the angular position of the loudspeakers. Figure 12 shows a picture of an exemplary setup in the IEM CUBE with loudspeakers, molleton, dummy head, and infrared cameras of the tracking system.

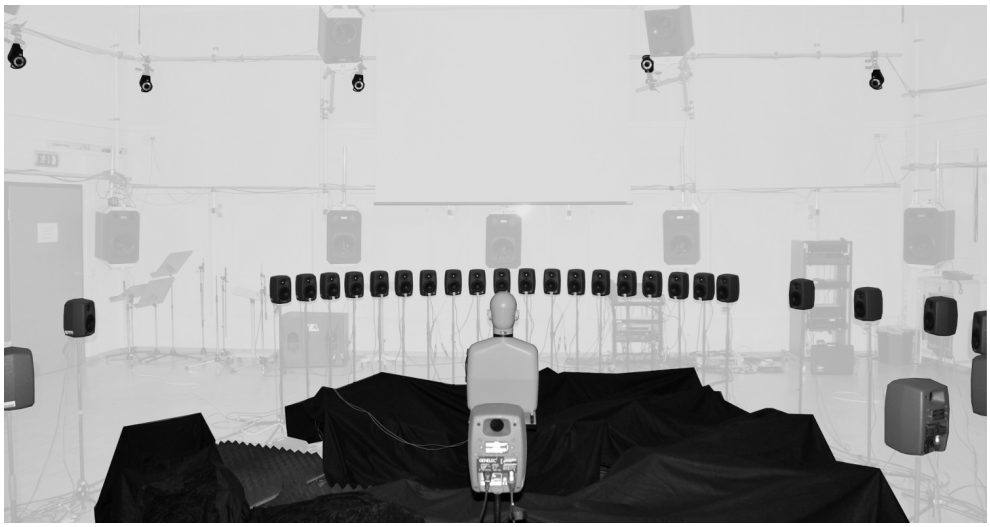


Figure 12: Picture of an exemplary experimental setup in the IEM CUBE: loudspeakers, molleton, dummy head, infrared cameras of the tracking system.

Using the molleton for suppressing floor reflections, the mean reverberation time of the CUBE is set to 470ms. Although the room is too large (10.3m×12m×4.8m) for surround reproduction according to ITU-R BS.1116-1 [ITU97], the reverberation time lies within the limits of the recommendation for most frequencies, cf. Figure 13.

The Genelec 8020 loudspeakers are arranged on a circle with a radius of 2.5m around the center of the CUBE. The loudspeakers are equalized in level and time delay at the central listening position, i.e. the origin of the arrangement. The omnidirectional critical distance of the setup is 2m. Assuming a directivity index of 6dB yields an effective critical distance of 2.8m. Thus, the central listening position lies within the effective critical distance. This is important for a good localization performance [Har83]. However, too little reverberation would produce prominent coloration as comb filters would be sparse and therefore more audible. For first order Ambisonics, the audio quality was found to be optimal when the loudspeakers are positioned approximately at the critical distance [SVP+09], which is the case for the experiments in this work.

The height of the loudspeakers (referred to halfway between woofer and tweeter, as recommended by the manufacturer [2]) was adjusted to 1.2m, which was also the

[2] `http://www.genelec.com/documents/other/acousticaxis.pdf`

ear height of the subjects and the dummy head.

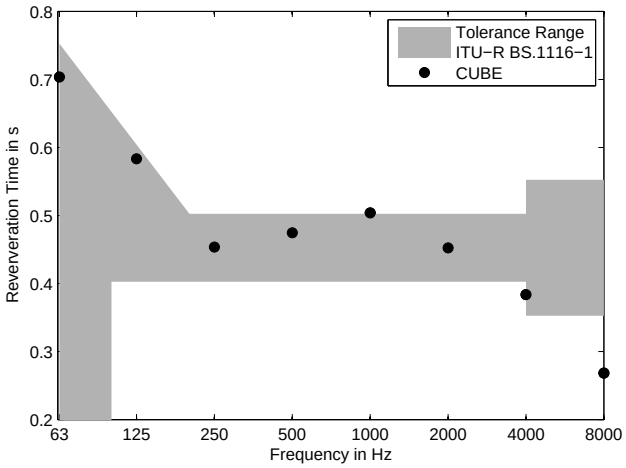


Figure 13: Reverberation time in the IEM CUBE and the ITU recommendation.

All listening experiments use pink noise with a level of 65dB(A) as stimulus. The broadband noise ensures good localization performance [Bla83]. The control of the entire experiments (except for the GUI of the coloration experiment in Section 5.3), as well as the creation of the loudspeaker signals used the open source software pure data[3] on a PC with an RME HDSPe MADI and RME M-16 DA D/A converters at a sample rate of 44.1kHz. The measurement of impulse responses employed exponential sine sweeps. All binaural measurements used a B&K 4128C dummy head, all other measurements used Schoeps CCM 8 figure-of-eight microphones and an NTI MM2210 omni-directional microphone. The measured impulse responses were convolved with A-weighted pink noise, in order to take into account human sound perception and the stimulus in the experiments.
For the measurement of the subjects' head movements and the pointed direction in the localization experiments, a VICON motion capture system equipped with 15 M-series cameras (infrared, 120 fps, covering the whole area within the loudspeaker arrangement) is employed. The resolution of the system is about 0.1mm and 0.1°, cf. Appendix A.

All subjects are part of an expert listening panel (ELP) that was recruited [SPH09], trained [FSH10], and monitored [FS12] in the context of the research project AAP. The ELP consists of 41 (14 female, 27 male) professional musicians and audio engineers or students in these fields. The age of the subjects was between 22 and 42 with an average of 28 years (in June 2013). The ELP also contributed to other experiments [FSLO12] of the research project that are not related to this work.

[3]freely available on `http://puredata.info/downloads`

2.3 Pointing Method for Localization Experiments

This section briefly presents the pointing method that is used for the localization experiments in Sections 3.2 and 3.3. More details can be found in the corresponding paper [FMSZ10].

There are several methods for measuring the perceived direction in localization experiments, which can be roughly divided in 4 categories: verbal methods [WK97], graphical methods [MFRB01], methods of adjustment [PK01, SMR09], and pointing methods. The latter have been proven to be most suitable in regard of intuitivity while retaining a high level of accuracy.

There are several different options for a subject to point towards a certain direction (an overview can be found in [See03]). Human gestures offer typical ways of doing so, e.g., looking into a certain direction, turning the head or even the whole body towards that direction, or using a combination of these gestures is intuitive. However pointing methods exhibit several restrictions. Pointing by eye and head gestures is restricted by the human physiology. Moreover, the tracking of eye movement is a technical challenge [SSJW10]. Results obtained by these pointing methods tend to be inaccurate due to a lack of proprioceptual decoupling [See03], and other issues [MLGM08]. However they can be improved by adding visual feedback, like pointing with a finger or an object held in one hand, especially in the vertical direction, where eye, head, and body movements are particularly limited [MLGM08]. Although the improvement reduces the bias, results deviate at a larger scale as the spatial resolution of human hearing [Bla83]. Other methods in which subjects use a joystick, trackball, or knob to control a laser point [See03, Lew98] achieve good proprioceptual decoupling and are accurate. Nevertheless, their handling may be counter-intuitive, and subjects tend to make too small movements with respect to the initially indicated direction [See03].

Figure 14: Toy-gun with reflective markers.

As pointing device, this work uses a toy-gun with laser and iron sights for optical aiming, cf. Figure 14. The position and orientation of the gun can be captured by a tracking system with a 6 degrees of freedom (DOF) sensor. The pointed direction is computed from the projection of the gun's direction on the surface of the surrounding hull or loudspeaker setup, cf. Figure 15. Thus the method works at any listening position within the range of the tracking system, for every nearly convex shaped hull or loudspeaker setup, and offers a full 3D directional coverage. The pointed direction can be referred to the position of the subjects head or to any other reference point, e.g., for spatial sound rendering system it is common to refer the pointed direction to the center of the loudspeaker arrangement.
As a model of the surrounding hull, either a convex hull or a simple ellipsoid can be used. In order to reduce measurement noise (e.g. the shaking of the hand and arm of the subject), the vectorized input data is temporally smoothed by filtering. The pointed direction can either be streamed continuously or stored as a single value at the trigger moment. In order to avoid deviations of the pointed direction during trigger actuation, the series of computed values is stored in a buffer. Reading out the buffer values that precede the trigger moment provides reasonable suppression of shaking.

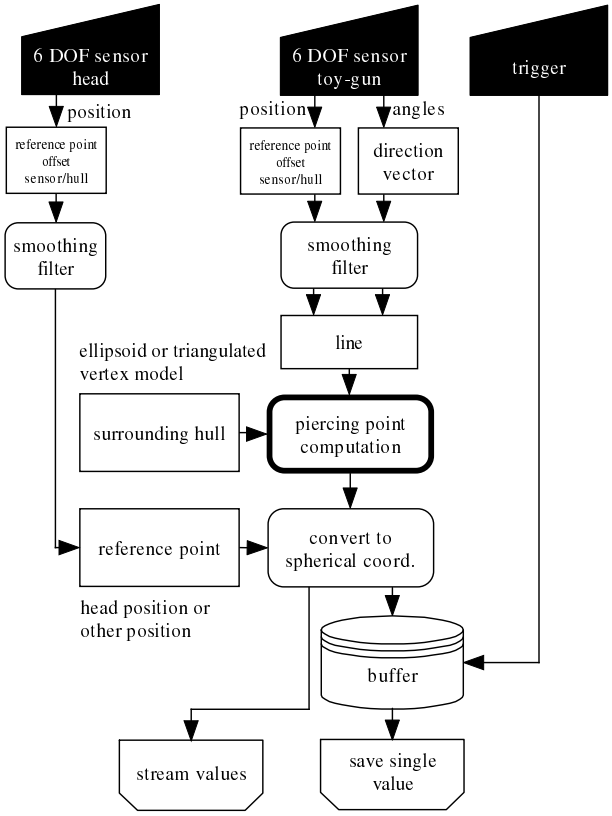


Figure 15: Computation of the pointed direction.

The trigger moment is captured by a game controller and transmitted to the computer. The buttons on this controller can be used for additional tasks, such as controlling the experiment sequence or evaluation of scale based attributes. A comparison of the measured accuracy when aiming at optical targets in 2 different rooms using 2 different tracking systems is presented in [FMSZ10].

Chapter III

PHANTOM SOURCE LOCALIZATION

The main goal of amplitude panning is the control of the phantom source location. So-called *panning laws* define the relation between the desired location and the gain of each loudspeaker. This chapter investigates how well these relations match with listening test results. It focuses on the localization of broadband pink noise stimuli, as these are known to yield good localization performance [Bla83].

Section 3.1 starts the investigation by a review of experimental localization results and panning laws from literature using pairwise panning between a loudspeaker pair at the standardized angles of $\pm 30°$ [ITU06]. The review also includes panning for non-standard arrangements and presents a suitable predictor for broadband stimuli.

The investigation is extended to amplitude panning on multiple loudspeakers using VBAP, MDAP, and Ambisonics with different weightings in Section 3.2. Although there are some studies about the localization of Ambisonics [BHL06, BDG$^+$07, FZ08] and a comparison to VBAP [KBBF07], there are no experiments that compare the localization of all above-mentioned panning methods. The section closes this gap by presenting a localization experiment at the central listening position. The listening test results are compared to different predictors, including a state-of-the-art binaural localization model.

Section 3.3 reveals what happens to localization at off-center listening positions, where additional level differences and time delays are present. The section also presents some studies about phantom source localization using vertical panning on three-dimensional loudspeaker arrangements. The section concludes with a discussion of causes for subjective differences in phantom source localization and presents a model of these differences.

3.1 Localization using 2 Loudspeakers

Pairwise panning laws have been determined either experimentally by listening tests [WT02] or were based on physical considerations [Pul97]. Although usually not considered in practice, they are known to depend on the spectrum of the source signal [WT02, PK01]. Most studies in literature focused on panning laws for the frontal standard stereo loudspeaker setup, i.e. loudspeakers at ±30° [ITU06]. There exist only few studies about the localization of phantom sources from lateral or dorsal directions or with other aperture angles between the loudspeakers [CDT+75, TP77, MWCQ99a, SMR09, SM10]. There are even fewer studies about panning laws for non-standard loudspeaker setups [Pul02] and there is no model to predict the subjective differences in the perceived directions.
This section reviews panning laws and experimental results for amplitude panning on standard ±30° loudspeaker setups and presents a prediction model for the localization of broadband phantom sources. This model is extended with the directivity of human hearing in order to model the localization of amplitude-panned phantom sources at any horizontal direction.

3.1.1 Localization of Frontal Phantom Sources

An overview and comparison of published listening experiment results can be found in [WT02]. For a given inter-channel level difference Δg in dB between the loudspeakers, the angular phantom source displacement is proportional to the aperture angle between the loudspeaker pair. Thus, Wittek defines the location of the phantom source in parts (%) of the loudspeaker aperture angle. He also found that the localization curves (perceived angle vs. inter-channel level difference) depend on the spectrum of the source signal. This explains the differences in the localization curves established by different laboratories, cf. black curves labeled with Wittek A…D in Figure 16. These are particularly inconsistent for panning directions close to the loudspeakers.

Tangent Law, Vector-Base Amplitude Panning, Velocity Vector

The first model for frontal pairwise amplitude panning was the *sine law* [Bau61] that predicts the phantom source position ϕ_s by the gains g_1 and g_2 of loudspeakers placed at $\pm\phi_L$

$$\frac{\sin\phi_s}{\sin\phi_L} = \frac{g_1 - g_2}{g_1 + g_2}. \tag{5}$$

It is based on a simple geometrical model of the head without accounting for the

propagation path around the head. In parallel, the *tangent law* was found [Lea59]

$$\frac{\tan\phi_s}{\tan\phi_L} = \frac{g_1 - g_2}{g_1 + g_2}. \tag{6}$$

It became more popular as it considered the propagation path around the head [BBE85] and thus seems to be more accurate. However in practice, the differences between both models have been shown to be negligible [Lea59].
Pulkki reformulated the tangent law in his work on *vector-base amplitude panning* (VBAP, cf. Section 2.1.1) [Pul97] to a vector form that allows also for non-symmetrical loudspeaker directions $\boldsymbol{\theta}_l = [\cos(\phi_l),\ \sin(\phi_l)]^{\mathrm{T}}$. The predicted direction $\boldsymbol{\theta}_s$ is calculated as a vector sum

$$\boldsymbol{\theta}_s = g_1\boldsymbol{\theta}_1 + g_2\boldsymbol{\theta}_2. \tag{7}$$

In order to achieve constant energy for all panning directions, the gains g_1 and g_2 have to be normalized with the overall energy by dividing with $\sqrt{g_1^2 + g_2^2}$.
The same linear summation of the weighted loudspeaker directions is used to calculate the *velocity vector* $\boldsymbol{r}_{\mathrm{V}}$ [Mak62, Ger92]

$$\boldsymbol{r}_{\mathrm{V}} = \frac{\sum_{l=1}^{\mathrm{L}} g_l\boldsymbol{\theta_l}}{\sum_{l=1}^{\mathrm{L}} g_l}. \tag{8}$$

The direction of this vector is assumed to correspond to the localization of low frequencies ($\leq$ 700Hz). Without limitation, the velocity vector can also be calculated for more than two simultaneously active loudspeakers.
As the tangent law, VBAP, and the velocity vector point towards the same direction, they are summarized and treated as one model in this article.

Energy Vector

Following the idea of the velocity vector, the *energy vector* $\boldsymbol{r}_{\mathrm{E}}$ [Ger92] has been defined as

$$\boldsymbol{r}_{\mathrm{E}} = \frac{\sum_{l=1}^{\mathrm{L}} g_l^2\boldsymbol{\theta_l}}{\sum_{l=1}^{\mathrm{L}} g_l^2} \tag{9}$$

This model assumes an energetic superposition of the loudspeaker signals and is expected to model the localization direction for higher frequencies or broadband signals. As the velocity vector, the energy vector has no restriction regarding the numbers active of loudspeakers. The magnitude of the energy vector can also used to describe spatial distribution of energy [Dan01] and the perceived width of phantom sources, cf. Section 4.1.2.

Sengpiel's Polynomial Model

Unlike the above-mentioned models, the model of Sengpiel [Sen12] is not based

on physical or geometrical assumptions. This model is based on experimentally determined localization curves and uses Lagrange interpolation to calculate a polynomial. The resulting relative phantom source shift $^{\Delta\phi_s}/_{\phi_L}$ is given in % of the angle ϕ_L for a symmetrical loudspeaker setup at $\pm\phi_L$ in dependence of the inter-channel level difference Δg in dB between the loudspeakers ($|\Delta g| \leq 18$dB, coefficients rounded to three fractional digits)

$$\frac{\Delta\phi_s}{\phi_l} = [1.729 \cdot 10^{-4}(\Delta g)^4 - 4.933 \cdot 10^{-3}(\Delta g)^3 - 0.149(\Delta g)^2 + 8.819\Delta g] \cdot 100\%. \tag{10}$$

Figure 16 shows the experimental results and the prediction models for pairwise amplitude panning in one diagram for a standard setup with loudspeakers at $\pm 30°$. In the experiments, the level difference Δg between the loudspeakers was varied from 0dB to 18dB. The same level differences are used to calculate and plot the model curves.

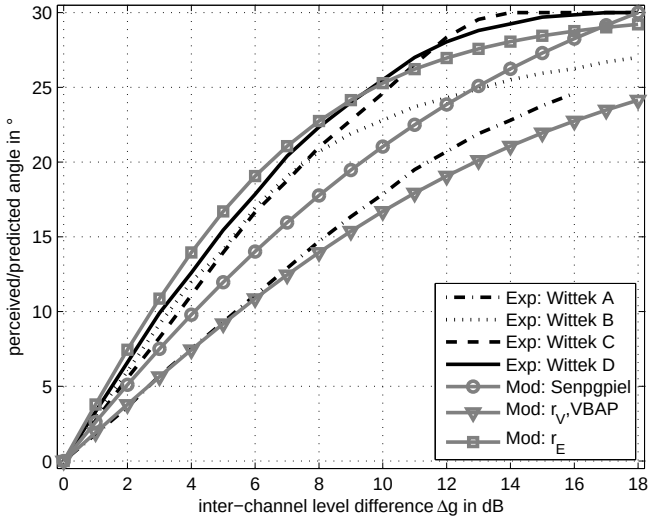


Figure 16: Localization curves (perceived/predicted angle vs. inter-channel level difference) for a standard $\pm 30°$ stereo loudspeaker setup. Experimental results (black) cited by [WT02] and model curves (gray).

The experimental results Wittek C and D and the energy vector model are similar. Up to 8dB these curves are also similar to Wittek B. Wittek A shows a smaller slope and is similar to the prediction of the velocity vector model. The polynomial model by Sengpiel lies in between all other curves. For a level difference of 18dB, the curves of Wittek C and D, as well as the polynomial model by Sengpiel, and the energy vector model yield a localization at the position of the loudspeaker. All other curves would need larger level differences to pan the phantom source on a loudspeaker.

The curve Wittek D was obtained with broadband maracas and claves sounds. Thus this curve seems to be the best starting point for a broadband predictor, and it matches the energy vector model.

3.1.2 Localization of Phantom Sources in the Horizontal Plane

Theile [TP77] conducted experiments with the standard aperture angle of 60° and rotated the loudspeaker pairs around the listener ($[10°, 70°]$, $[30°, 90°]$, $[50°, 110°]$, and $[60°, 120°]$). For a level difference Δg of 0dB, he found out that the localized position of the phantom source is not longer exactly in the middle between the loudspeakers. The localization rather seems to be dominated by the most frontal loudspeaker. This tendency increases the more lateral the loudspeaker pair is. Additionally, for this situation localization flips to the loudspeaker positions for level differences $|\Delta g| \geq 6$dB. The localization curves are governed by a steep slope and strong subjective variation in the perceived direction for level differences $|\Delta g| < 6$dB.
Similar experiments were conducted by Pulkki [Pul02] for ±30°, $[0°, 60°]$, and $[30°, 90°]$, and yielded comparable results. His experiments aimed at finding a compensation for the phantom source displacement.
The same tendencies were found in the experiments from Cabot [CDT+75] for quadrophonic setups (±45° and ±135°), Martin [MWCQ99a] for 3/2 surround setups ($[0°, 30°]$, $[30°, 120°]$, and ±120°) and Simon [SMR09] for an aperture angle of 45° ($[0°, 45°]$, $[45°, 90°]$, $[90°, 135°]$, and $[135°, 180°]$). The latter author compared his results to the localization that is predicted by VBAP and states that VBAP is not suitable for the prediction of lateral phantom sources.

Non-Unitary Vector-Base Amplitude Panning

Pulkki conducted experiments to study the displacement of phantom sources for loudspeaker positions that are not symmetrical to the median plane [Pul02]. He used three different loudspeaker configurations (±30°, $[0°, 60°]$, and $[30°, 90°]$). Consistently with above, the experiments revealed that there is a bias towards the median plane. To compensate for the displacement by attenuating the loudspeaker that is closer to the median plane, Pulkki introduced a non-unitary base for VBAP. This model was designed for loudspeaker positions in front of the inter-aural axis and is thus not suitable for the prediction of phantom source localization using all horizontal directions.

Weighted Energy Vector

Direction-dependent weighting $w(\boldsymbol{\theta_l})$ can also be applied to the energy vector

[Ger92] by simply weighting of the loudspeaker gains [FMS11]. This yields the weighted energy vector $\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}}$

$$\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}} = \frac{\sum_{l=1}^{\mathrm{L}} (g_l w(\boldsymbol{\theta_l}))^2 \boldsymbol{\theta_l}}{\sum_{l=1}^{\mathrm{L}} (g_l w(\boldsymbol{\theta_l}))^2}. \tag{11}$$

Appropriate weights have to consider the dependence of human loudness perception on different directions. Thus it is useful to apply the directivity of human hearing as directional weighting.

In the late fifties, the directivity of a single human ear has been measured with microphones [JV59]. Later on, Jahn conducted experiments on how a single value for the perceived binaural loudness could be determined from the sound pressure of both ears $p_{left}(\phi)$ and $p_{right}(\phi)$ [Jah63]. He ended up with the following summation formula for the directional loudness or directivity $D(\phi)$

$$D(\phi) = \left(\frac{|p_{left}(\phi)|^k + |p_{right}(\phi)|^k}{|p_{left}(0)|^k + |p_{right}(0)|^k} \right)^{\frac{1}{k}}. \tag{12}$$

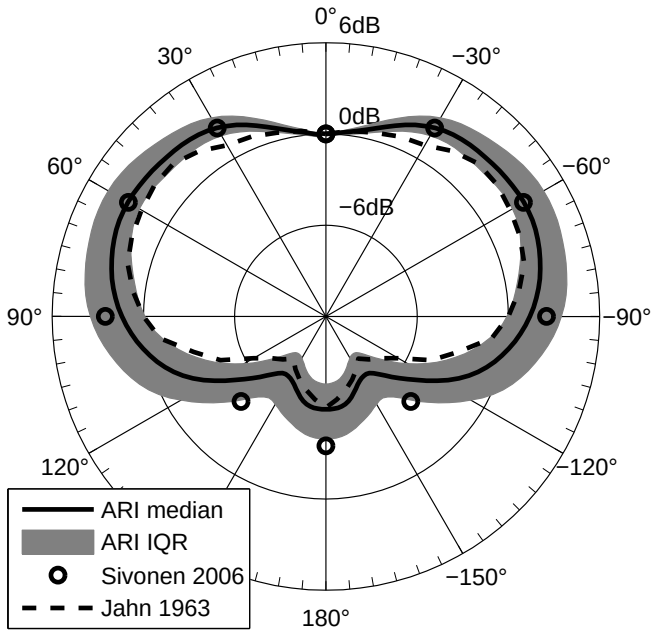
He found that this formula is valid for arbitrary frequencies. However, the exponent k depends on the loudness level of the sound: $k = 2.1$ for levels close to the absolute hearing threshold and $k = 1.77$ for higher levels (50-70 phon). A later work found that an exponent of $k = 2$ yields the best correlation to listening test results, i.e. an energetic superposition of both ear signals [SE06]. This value is also used in this work.

Using Eq. (12), the directivity of human hearing can be computed directly from head-related transfer functions (HRTFs). In order to have a good estimate of the average directivity, this work employs an HRTF database from the Acoustic Research Institute (ARI) in Vienna with 66 normal hearing subjects (now 85) [1]. For each subject, the ARI database includes 1550 measured directions. In the horizontal plane, the HRTFs were measured with a resolution of 2.5° within the azimuth range of ±45° and with a resolution of 5° outside this range. Figure 17(a) shows the median directivity and the corresponding interquartile range (IQR) in the horizontal plane computed from the ARI database for a third-octave band around 5kHz. For comparison to experimentally determined results from the literature, the directivities of a third-octave band around 5kHz from [SE06] and around 4.5kHz from [Jah63] are also shown in the plot. All directivities yield similar values.

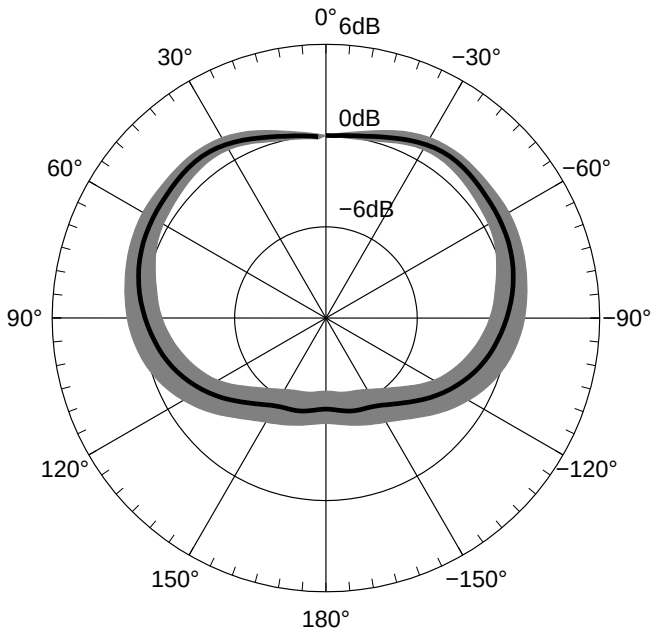
For this work, the broadband directivity using A-weighted pink noise was computed from the ARI database and is shown in Figure 17(b). The median of this

[1] freely available on `http://www.kfs.oeaw.ac.at/index.php?lang=en`

broadband directivity is employed as direction-dependent weighting $w(\boldsymbol{\theta_l})$ for $\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}}$ with $\boldsymbol{\theta}_l = [\cos(\phi_l),\ \sin(\phi_l)]^{\mathrm{T}}$.



(a) third-octave band: computed from ARI database (5kHz center frequency), subjective results from [SE06] (5kHz), and computed data from [Jah63] (4.5kHz)



(b) broadband (A-weighted pink noise): computed from ARI database

Figure 17: Polar pattern of horizontal directivity of human hearing, different frequency bands, experimental and computed data.

The following paragraphs compare selected experimental results and models of localization curves in the horizontal plane. The experimental results are from Theile [TP77] ($[10°, 70°]$, $[30°, 90°]$, $[50°, 110°]$, and $[60°, 120°]$) using noise impulses (600Hz − 10kHz), Martin [MWCQ99a] ($[0°, 30°]$, $[30°, 120°]$, and $\pm 120°$) using continuous female speech, and Simon [SMR09] ($[0°, 45°]$, $[45°, 90°]$, $[90°, 135°]$, and $[135°, 180°]$) representing average results for female speech, cello, bongos, and pink noise.

The experimentally determined localization curves are compared to the VBAP / velocity vector $\boldsymbol{r}_\mathrm{V}$ and the energy vector $\boldsymbol{r}_\mathrm{E}$ model. In addition, the weighted energy vector $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$ is included in the comparison.

Tables 2 and 3 compare the deviation of the different prediction models from the mean/median values of the experimental results. Table 2 shows the absolute deviation of the predicted phantom source locations from the experimental results averaged over all level differences, and Table 3 shows the deviation for a level difference of $\Delta g = 0$dB, i.e., when both loudspeakers play at the same level. Both deviation measures are given in % of the loudspeaker aperture angle in Tables 2 and 3.

experiment	**positions** ϕ_l		$\boldsymbol{r}_\mathrm{V}$	$\boldsymbol{r}_\mathrm{E}$	$\boldsymbol{r}_\mathrm{E}^\mathrm{w}$
Theile, 1977 [TP77]	10°	70°	6.5%	4.8%	5.5%
Theile, 1977 [TP77]	30°	90°	9.1%	5.3%	3.4%
Theile, 1977 [TP77]	50°	110°	15.8%	8.0%	4.2%
Theile, 1977 [TP77]	60°	120°	18.2%	7.7%	8.5%
Martin, 1999 [MWCQ99a]	0°	30°	9.6%	6.1%	4.9%
Martin, 1999 [MWCQ99a]	30°	120°	9.8%	6.8%	5.7%
Martin, 1999 [MWCQ99a]	−120°	120°	21.4%	11.8%	11.8%
Simon, 2009 [SMR09]	0°	45°	7.5%	4.9%	6.4%
Simon, 2009 [SMR09]	45°	90°	10.2%	2.9%	5.9%
Simon, 2009 [SMR09]	90°	135°	16.9%	14.6%	5.1%
Simon, 2009 [SMR09]	135°	180°	10.5%	10.1%	6.0%
mean deviation:			12.3%	7.5%	6.1%

Table 2: Average absolute deviation of the models from experimental mean/median results. Used model are: VBAP / velocity vector $\boldsymbol{r}_\mathrm{V}$, energy vector $\boldsymbol{r}_\mathrm{E}$, and directivity-weighted energy vector $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$. Deviations are given in % of the loudspeaker aperture angle.

Regarding the average absolute deviation, cf. Table 2, both energy vector models perform significantly better than the velocity vector model. This finding supports the suitability of the energy vector for broadband signals. The deviations of the energy vector and the directivity-weighted energy vector are both similar except for lateral phantom sources for which the weighting by the directivity improves the model predictions.

experiment	**positions** ϕ_l		$\boldsymbol{r}_\mathrm{V}$	$\boldsymbol{r}_\mathrm{E}$	$\boldsymbol{r}_\mathrm{E}^\mathrm{w}$
Theile, 1977 [TP77]	10°	70°	2.2%	2.2%	8.7%
Theile, 1977 [TP77]	30°	90°	13.3%	13.3%	2.1%
Theile, 1977 [TP77]	50°	110°	24.4%	24.4%	1.2%
Theile, 1977 [TP77]	60°	120°	30.0%	30.0%	0.7%
Martin, 1999 [MWCQ99a]	0°	30°	6.7%	6.7%	1.9%
Martin, 1999 [MWCQ99a]	30°	120°	19.4%	19.4%	30.7%
Martin, 1999 [MWCQ99a]	−120°	120°	1.1%	1.1%	1.1%
Simon, 2009 [SMR09]	0°	45°	7.8%	7.8%	16.0%
Simon, 2009 [SMR09]	45°	90°	3.7%	3.7%	12.6%
Simon, 2009 [SMR09]	90°	135°	35.6%	35.6%	12.7%
Simon, 2009 [SMR09]	135°	180°	21.7%	21.7%	10.7%
	mean deviation:		15.1%	15.1%	8.9%

Table 3: Absolute deviation of the models from experimental mean/median results for a level difference of $\Delta\mathrm{dB}(g) = 0\mathrm{dB}$. Used model are: VBAP / velocity vector $\boldsymbol{r}_\mathrm{V}$, energy vector $\boldsymbol{r}_\mathrm{E}$, and directivity-weighted energy vector $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$. Deviations are given in % of the loudspeaker aperture angle.

For a level difference of $\Delta g = 0\mathrm{dB}$, cf. Table 3, the velocity vector and the (unweighted) energy vector model yield the same results, because there is no difference between linear and energetic superposition in this case. The weighted energy vector mostly achieves smaller deviations, particularly whenever at least one loudspeaker is positioned behind the inter-aural axis. An exception of this finding appears in the experiments of Martin [MWCQ99a] for loudspeakers at $[30°, 120°]$, but the subjective variation in the data of this case was also huge.

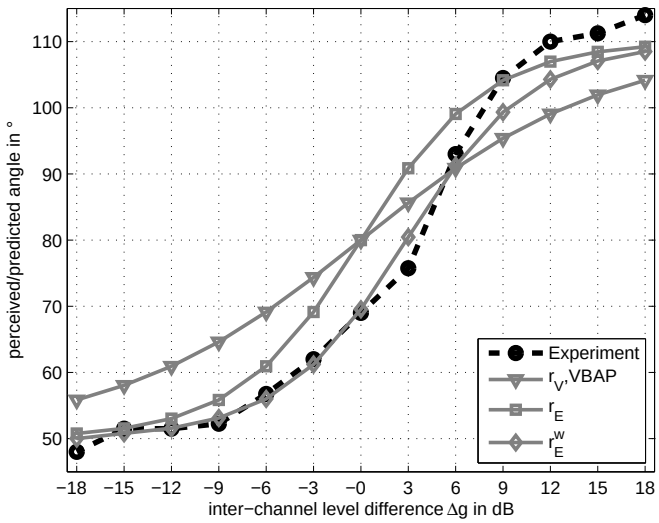


Figure 18: Localization curves (perceived/predicted angle vs. inter-channel level difference) for a $[50°, 110°]$ loudspeaker setup from Theile [TP77].

Figure 18 exemplary shows the comparison of the localization curves for the three models and the experimental results from Theile [TP77] for loudspeakers at $[50°, 110°]$.
Summarizing, the direction of the directivity weighted energy vector $\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}}$ seems to be the best predictor for the median localization of the reviewed experiments. In addition, considering the subjective differences in the phantom source localization model seems to be a fruitful extension.

3.1.3 Discussion

This section reviewed experimental results and models for the localization in pairwise amplitude panning on standard $\pm 30°$ loudspeaker setups. It has been shown that the direction of the energy vector is a good predictor for the localization of broadband phantom sources.
Experimental results indicate that frontal panning models are not ideal for phantom sources at side and rear positions. Basically, if there are no level differences, the phantom sources are not localized exactly in the middle between the two loudspeakers. In this case, the experimental results yield a large subjective variation. Furthermore, the localization sticks to the loudspeaker positions. The energy vector prediction model is extended for pairwise amplitude panning in the entire horizontal plane by incorporating the median directivity of human hearing that is computed from an HRTF database.

3.2 Localization using Amplitude Panning with Multiple Loudspeakers

This section expands the previous study to the panning methods MDAP, Ambi$^{\text{RE}}$, and Ambi$^{\text{RV}}$ by a listening experiment and investigates predictors for the experimental results. For this purpose, two additional predictors are introduced. The section addresses the questions of how close the localized directions of the different panning methods are compared to the desired directions, and how good the predictors match with the experimental results.

3.2.1 Additional Localization Predictors

Binaural Localization Model

In order to predict localization, binaural models are common that use dummy head recordings as input. This work employs a localization model after Lindemann [Lin86a, Lin86b] which is part of the Auditory Modeling Toolbox[2]. It divides the binaural input signals into 36 frequency bands (center frequencies from 164-16935Hz) with a spacing of 1 ERB (equivalent rectangular bandwidth) [MPG90]. The auditory nerve is modeled by a half-wave rectifier and a low-pass filter at 800Hz. In each band, the inter-aural level-difference (ILD) is considered by monaural detectors and contra-lateral inhibition that weight the delayed signals for the computation of the inter-aural cross correlation function (IACF). This yields a displacement of the peaks in the IACF. The inter-aural time-difference (ITD) is then computed as the centroid of the IACF [Jef48], which delivers one ITD value for each frequency band. In order to study the temporal behavior of localization, running cross correlation was introduced in [Hes07].

Figure 19 shows the ITD values of real sound sources for different frequencies bands and incidence angles computed by the binaural model. Assuming summing localization, phantom sources at the same positions would result in similar values. Pulkki showed that this is at least the case for frequencies $\leq$ 1.1kHz [PK01]. Therefore, this work employs the binaural model as predictor for phantom source localization. This model is also used for the prediction of localization for other spatial sound rendering methods, e.g. wave field synthesis [WRGS13].

As described above, the model yields one ITD value for each frequency band. Within each frequency band, the ITD value of the phantom source is compared to the values of real sources in a lookup table, cf. Figure 19. The best matching ITD is selected (using linear interpolation between neighboring values) and the corresponding angle is regarded as the angle of the phantom source for the present

[2] freely available on `amtoolbox.sourceforge.net/`

frequency band. A single angle as prediction result is achieved by the median value of the angles for all frequency bands. As the presented predictor does not apply frequency-dependent weighting of the ITD values, binaural room impulse responses can be directly used instead of recorded stimuli. The model uses the first 80ms of the impulse responses.

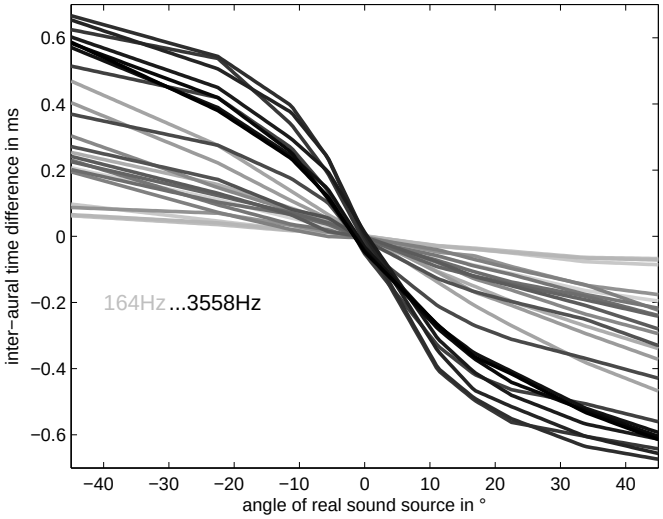


Figure 19: Frequency-dependent ITDs for real sources between ±45°.

The model cannot distinguish between front and back, thus the ITD values in the lookup table were limited to the directions between ±45°, between which the conditions of the experiment in Section 3.2 lie. The best fit to the median experimental results has been achieved when using 21 frequency bands covering the range from 164Hz-3558Hz; a fact that supports the importance of the ITD for higher frequencies [MM02].

Intensity Vector

Sound intensity is a physical measure of the directional sound power flow and can thus be used to determine the direction where sound is coming from [Mer07]. The intensity $\boldsymbol{I}$ is computed from the scalar sound pressure p and the vectorial particle velocity $\boldsymbol{v}$ as $\boldsymbol{I} = p\boldsymbol{v}$ [Hey86]. The sound pressure can be measured by an omni-directional microphone. The particle velocity is typically not measured directly but by the pressure gradient in the directions of the x-, y-, and z-axis. In the two-dimensional case considered here, this is done by two figure-of-eight microphones, each one aligned to the axis of the coordinate system labeled with v_x and v_y corresponding to their alignment. Here p, v_x, and v_y are computed as the convolution of measured impulse responses with A-weighted pink noise.
As a predictor for sound source directions, it is not suitable to compute the instantaneous direction of the intensity vector for each sample separately (each 22.7μs at

a sample rate of 44.1kHz), but rather as a temporal average within a certain time window. The time window was set to 80ms (S=3528 samples), which corresponds to the binaural localization model. The components $\overline{I_x}$ and $\overline{I_y}$ of the temporally averaged intensity vector $\overline{\boldsymbol{I}} = \left[\overline{I_x}\ \overline{I_y}\right]$ are computed as

$$\overline{I_x} = \sum_{s=1}^{\mathrm{S}} p(s)v_x(s) \qquad \text{and} \qquad \overline{I_y} = \sum_{s=1}^{\mathrm{S}} p(s)v_y(s). \tag{13}$$

The direction of the intensity vector is calculated as $\arctan(\overline{I_y}, \overline{I_x})$ and is equal to the direction of the velocity vector under free-field conditions. In the literature, the intensity is often computed in the frequency domain [Pul07] which also results in temporal smoothing and similar results.

3.2.2 Experiment

The listening experiment evaluates the localization of phantom sources created by VBAP, MDAP, Ambi$^{\mathrm{rE}}$, and Ambi$^{\mathrm{rV}}$ on a regular ring of 8 loudspeakers, i.e. loudspeakers with an aperture angle of $\Delta\phi_L = 45°$. Both Ambisonics variants use a maximum order of 3 and MDAP is applied with B = 10 panning directions uniformly distributed within a spread of $\phi_{\mathrm{MDAP}} = 22.5°$. The influence of different Ambisonics orders has already been studied [FZ08] and is not part of this experiment. Figure 20 shows the experimental setup with additional inactive but visible loudspeakers placed in 5° steps in and around the angular range of the target directions.

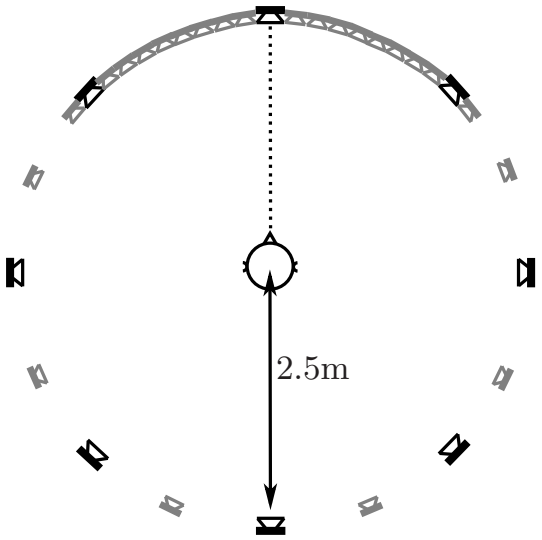


Figure 20: Experimental setup using 8 (black) loudspeakers; the smaller gray loudspeakers were visible but inactive.

All panning methods were evaluated for 9 directions between 0° and −45° (to the right). Each of the 36 = 9 (directions) × 4 (panning-methods) conditions was

repeated once. The stimuli were 3 pink noise bursts, each with 100ms fade-in, 200ms full scale level (65dB(A)), 100ms fade-out, and 200ms silence before the next fade-in. The stimuli could be repeated at will by the subjects. The perceived direction was assessed by the pointing method that is described in Section 2.3. The results for the central listening position are presented here. For results at an off-center position see Section 3.3.1.

There were 14 subjects participating in the experiment. As each condition was evaluated twice, 28 answers are available for each condition. Figure 21 shows the median values and the corresponding 95% confidence intervals for the 9 different panning angles using VBAP at the central listening position. In comparison to the ideal localization curve (perceived angle = panning angle), the perceived angles tends toward the loudspeakers. This tendency is known from the review of the panning laws in Section 3.1 and is also reproduced by the weighted energy vector. The binaural model predicts this tendency only partly. However all predictors deviate less than 2.8° from the median of the experimental results on average. For a detailed comparison of the predictors see Table 6.

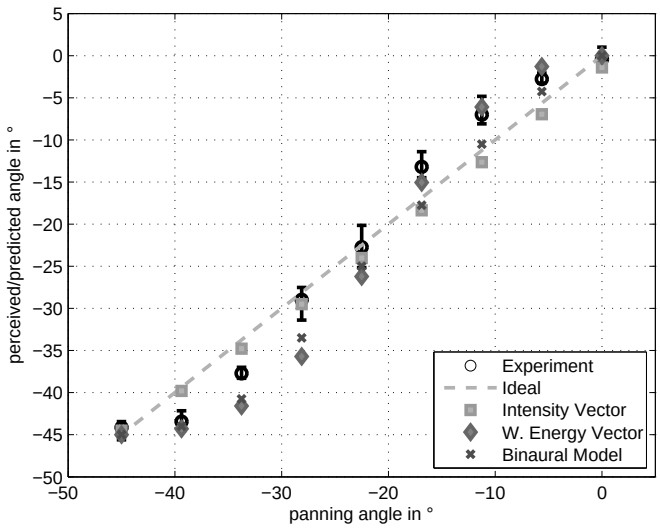


Figure 21: Median values and corresponding 95% confidence intervals of the perceived angles for different panning angles within $[0° \; -45°]$ using VBAP at the central listening position, as well as ideal localization curve and predictions by intensity vector, weighted energy vector, and binaural model.

Figure 22 presents the results for MDAP. In comparison to VBAP, the median perceived angle is closer to the ideal localization curve. The predictors perform similar as before, with a slight improvement of the intensity vector and a slight impairment of the binaural model.

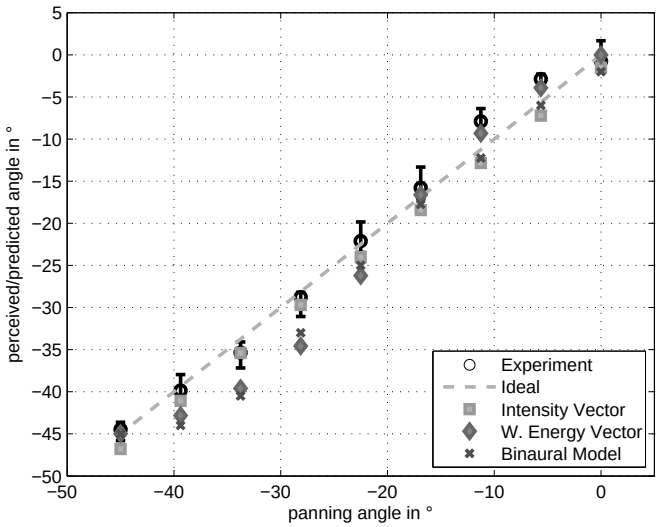


Figure 22: Median values and corresponding 95% confidence intervals of the perceived angles for different panning angles within $[0° \; -45°]$ using MDAP at the central listening position, as well as ideal localization curve and predictions by intensity vector, weighted energy vector, and binaural model.

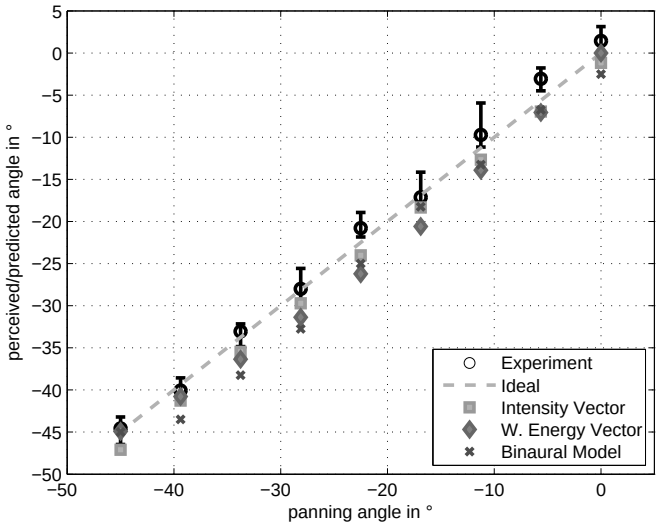


Figure 23: Median values and corresponding 95% confidence intervals of the perceived angles for different panning angles within $[0° \; -45°]$ using Ambi$^{\text{rE}}$ at the central listening position, as well as ideal localization curve and predictions by intensity vector, weighted energy vector, and binaural model.

The median perceived angles are even closer to the ideal localization curve when using Ambi$^{\text{rE}}$, cf. Figure 23. In this case, all predictors perform slightly worse compared to MDAP, with the intensity vector still being the best predictor. For Ambi$^{\text{rV}}$, cf. Figure 24, the binaural model poorly predicts the localization of the lateral panning angles. The weighted energy vector performs better and the intensity vector is the best predictor. Compared to VBAP, the match between the median experimental results and the ideal localization curve is better for Ambi$^{\text{rV}}$.

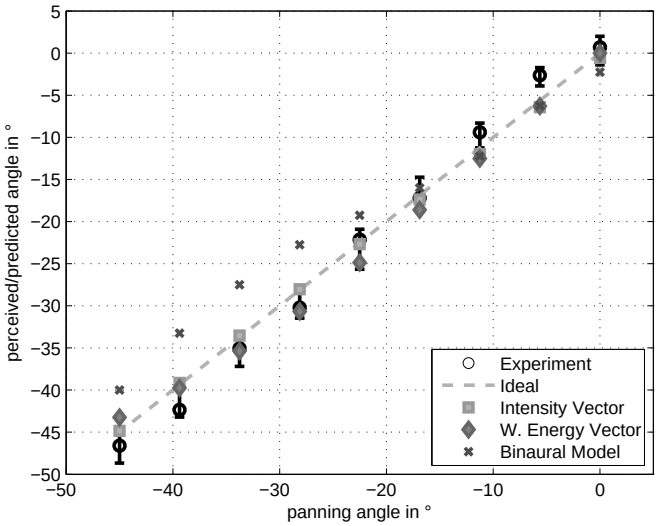


Figure 24: Median values and corresponding 95% confidence intervals of the perceived angles for different panning angles within $[0° \; -45°]$ using Ambi$^{\text{rV}}$ at the central listening position, as well as ideal localization curve and predictions by intensity vector, weighted energy vector, and binaural model.

3.2.3 Discussion

Table 4 compares the median deviation of the different panning methods from the ideal localization curve, i.e., how much the perceived angle deviated from the panning angle. The angles deviate most for VBAP, in fact more than two times as much as for Ambi$^{\text{rE}}$. This agrees with the findings in Section 3.1 that VBAP applies the tangent law, which is assumed for the localization of low-frequency stimuli and obviously does not perfectly fit for broadband stimuli. The angular match is best for Ambi$^{\text{rE}}$, followed by MDAP, Ambi$^{\text{rV}}$, and VBAP with the poorest match.

VBAP	**MDAP**	**Ambi$^{\text{rE}}$**	**Ambi$^{\text{rV}}$**
2.35°	1.28°	1.05°	1.58°

Table 4: Average absolute deviation of median experimental results from ideal localization curve for different panning methods.

Despite this ranking, VBAP yields the narrowest confidence intervals in the experimental results, cf. Table 5. This is mainly the case for panning angles close to the loudspeakers, which provides narrow and accurate localization in VBAP in comparison to phantom sources created by other methods. Section 3.3.2 discusses possible reasons for subjective differences in phantom source localization. Still, the confidence intervals for other panning methods are larger by only 0.8°.

VBAP	**MDAP**	**Ambi^{r_E}**	**Ambi^{r_V}**
2.75°	3.06°	3.48°	3.53°

Table 5: Mean 95% confidence intervals of experimental results for different panning methods.

Table 6 compares the deviation of the predictions from the median experimental results. Interestingly, the worst prediction with a deviations between 2.35° and 4.92° is achieved by the binaural model, which is at the same time the most complex predictor. The intensity vector is the best predictor for MDAP, Ambi^{r_E}, and Ambi^{r_E} and it matches the experimental results with a deviation between 1.89° and 2.75°. On average, the weighted energy vector performs nearly equally good as the intensity vector, with an average deviation between 2.04° and 2.92°. This is remarkable as this predictor is the simplest and does not require measurements.

	VBAP	**MDAP**	**Ambi^{r_E}**	**Ambi^{r_V}**
Binaural Model	2.35°	3.07°	3.37°	4.92°
Intensity Vector	2.75°	2.12°	2.41°	1.89°
Weighted Energy Vector	2.27°	2.40°	2.92°	2.04°

Table 6: Average absolute deviation of predictions from median experimental results for different panning methods.

The predictions by the intensity vector are very close to the ideal localization curve. This is obvious because the direction of the intensity vector is the measured counterpart of the velocity vector. In turn, the direction of the velocity vector is identical to the panning direction for the evaluated panning methods.

3.3 Further Studies on Phantom Source Localization

3.3.1 Off-Center Listening Positions

The results from the experiment in Section 3.2 showed that localization works similarly good for all studied panning methods when the subjects are seated in the center of the loudspeaker arrangement. For off-center listening positions, the distances to the listener are no longer the same for all loudspeakers. The different distances introduce additional time and level differences that modify localization. The results in [FZ08, KBBF07] show that the localization is obviously dominated by the nearest active loudspeaker. As, for higher Ambisonics orders, the energy is concentrated around the panning direction, the localization errors are smaller.
This section investigates the localization for one off-center listening position, cf. Figure 25. The listening experiment compares VBAP, MDAP, Ambi$^{\text{rE}}$, and Ambi$^{\text{rV}}$ for panning directions between 0° and −45°. The entire experiment (setup, conditions, number of repetitions, pointing method, and subjects) is identical to the experiment in Section 3.2 except for the listening position. The subjects were facing the loudspeaker at 0°.

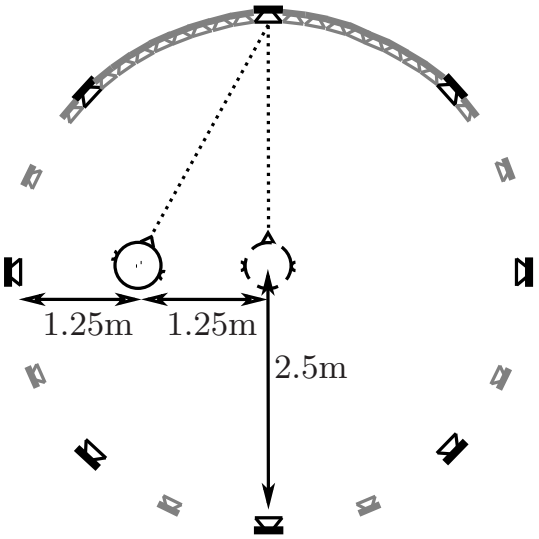
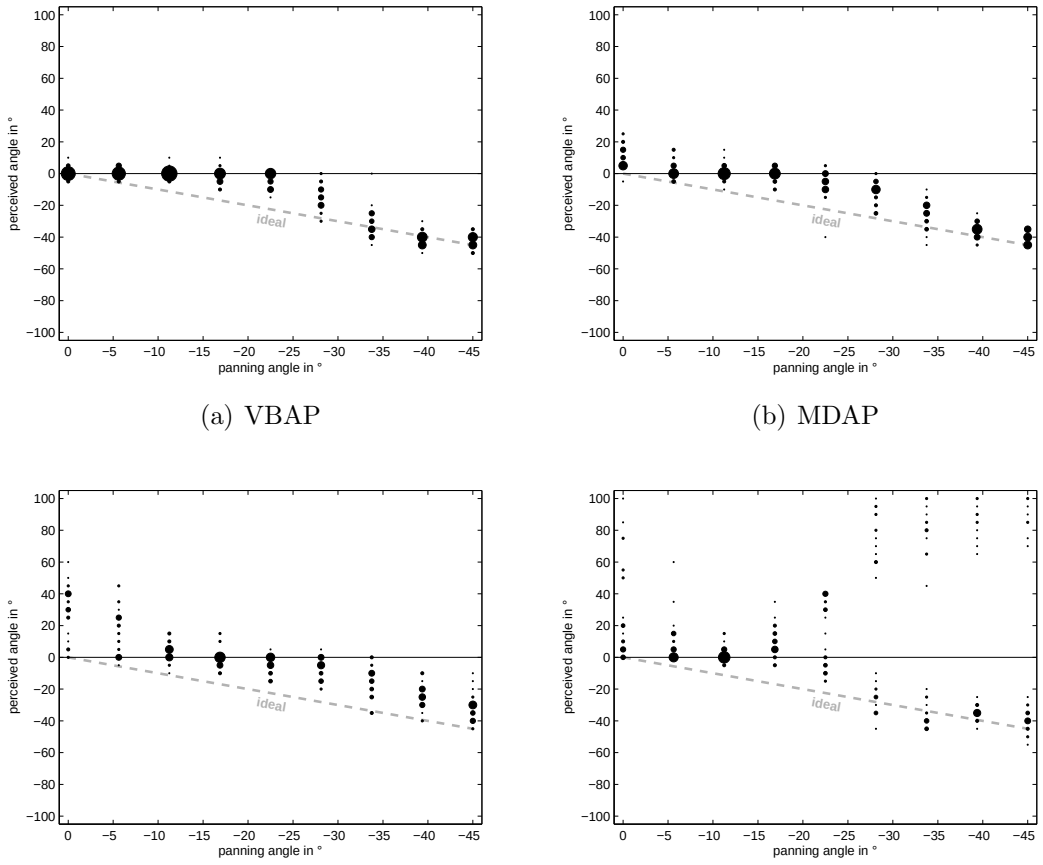


Figure 25: Experimental setup using 8 (black) loudspeakers; the smaller gray loudspeakers were visible but inactive, dashed head shows central listening position as a reference.

Figures 26(a) to 26(d) show the resulting histograms of the perceived angles for different panning angles. The dashed gray line shows the ideal localization curve (perceived angle = panning angle) as a reference. For VBAP, the perceived angles stay within the investigated panning range of $[0° \ -45°]$, resulting in an average absolute deviation of 9°. However, the localization sticks to the loudspeakers,

especially to the closer loudspeaker at 0°, which agrees with the precedence effect [Haa72, WNR73, HOT+07]. This tendency can also be found when using MDAP. For MDAP, some phantom sources are even perceived from outside the panning range, i.e. at directions left from the 0° direction. The average absolute deviation using MDAP is 11°. The localization shift to the left is even stronger for Ambi$^{\mathrm{r_E}}$ yielding an average absolute deviation of 18°. Despite the offset, the localization curve follows the desired tendency and is more or less parallel to the ideal curve. For VBAP, MDAP, and Ambi$^{\mathrm{r_E}}$, the average absolute deviation of the experimental results stays smaller than half the loudspeaker aperture angle. At the considered off-center listening position, it can be assumed that the deviation is proportional to the loudspeaker aperture angle and the Ambisonics order.



(a) VBAP (b) MDAP

(c) Ambi$^{\mathrm{r_E}}$ (d) Ambi$^{\mathrm{r_V}}$

Figure 26: Histograms of the perceived angles at the off-center listening position for different panning angles using VBAP, MDAP, Ambi$^{\mathrm{r_E}}$, and Ambi$^{\mathrm{r_V}}$. The size of the circles indicate the relative frequency of each perceived direction.

This does not hold true for Ambi$^{\mathrm{r_V}}$. The subjects perceived a second auditory event that splits from the one at the desired direction. This second event is located at about 90° to the left, where the loudspeaker is located that is closest

to the listener. This source splitting does only occur for Ambi$^{\text{rv}}$ as this panning method spreads the signal to loudspeakers distant to the desired auditory event, due to its strong side lobes. The source splitting is also known from headphone experiments [BS13] and has an effect on the perceived coloration change for moving phantom sources, cf. Section 5.3.
The prediction of the experimental results would require models that do not only incorporate level and time differences, but also a probabilistic assessment of directions instead of one single angle in order to explain the source splitting.

3.3.2 Subjective Differences in Horizontal Phantom Source Localization

In Section 3.1.2, the large variation in the subjective results of lateral phantom sources with small level differences has already been mentioned. This variation could be due to intrasubjective or intersubjective variation. Intrasubjective variation means that the same subject gives different answers for the same repeated question, whereas intersubjective describes the differences between the answers of different subjects. Theile [TP77] mentioned intersubjective variations and Pulkki [Pul01a] showed results for three-dimensional VBAP which suggest that intersubjective variations are larger than intrasubjective variations.
This section investigates the source of the subjective differences in the localization of phantom sources. This is done by additional conditions in the experiment of Section 3.2 using phantom sources created by loudspeaker pairs with a level difference of $\Delta g = 0$dB. The loudspeaker pairs were located at $[0°, 45°]$ and $[90°, 135°]$. For each subject, the two conditions were repeated six times randomly distributed over the entire experiment.

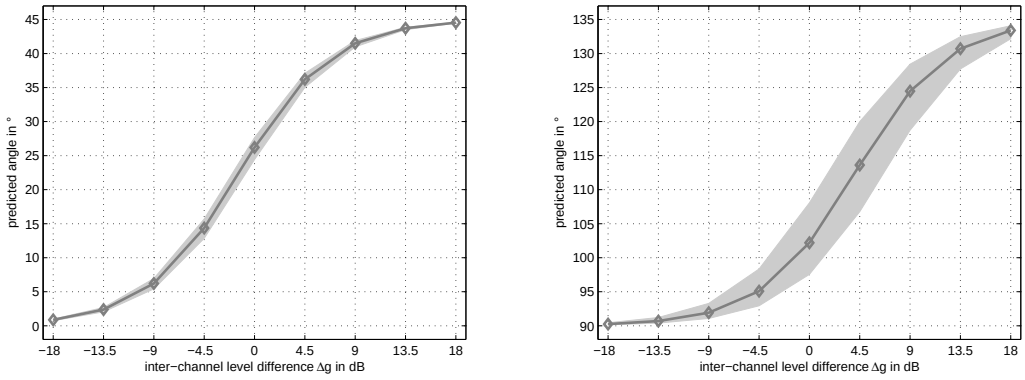
loudspeakers	**overall**	**intersubjective**	**intrasubjective**
$[0°, 45°]$	2.09°	1.13°	1.01°
$[90°, 135°]$	5.93°	5.66°	2.82°

Table 7: 95% confidence intervals of the median localized directions for loudspeaker pairs with $\Delta g = 0$dB at $[0°, 45°]$ and $[90°, 135°]$: overall, intersubjective, and intrasubjective intervals

Table 7 compares the overall, intersubjective, and intrasubjective confidence intervals for the median localized directions. The overall confidence intervals are computed for the $84 = 14$ (subjects) $\times$ 6 (repetitions) answers representing the overall variation caused by intersubjective and intrasubjective differences for each loudspeaker pair. The intersubjective confidence interval is calculated after replacing each subject's answers by her/his median answer for each loudspeaker pair.

Vice versa, the intrasubjective confidence is calculated after suppressing the intersubjective variation of each subject by subtracting the subject's median answer. The intersubjective variation is a bit greater than the intrasubjective variation for the loudspeaker pair at $[0°, 45°]$. An analysis of variance (ANOVA) shows a significant effect of the subjects ($p \ll 0.001$), but no effect of the repetitions ($p = 0.17$). For the loudspeaker pair at $[90°, 135°]$, the intersubjective variation is twice as big as the intrasubjective variation. In this case, nearly the whole overall variation is caused by intersubjective variation. Again, the effect of the subjects is significant ($p \ll 0.001$), but the effect of the repetitions is not ($p = 0.75$). This finding shows that the subjective differences in the localization is dominated by differences between the subjects. The results also indicate these differences being greater for the lateral loudspeaker pair.

In order to model the intersubjective variation, it is necessary to reveal the cause of this variation, i.e., what the difference between the subjects is. It is assumed that the HRTFs of different subjects differ and thus the directivity of subject's hearing. Instead of using only the median directivity for the weighted energy vector, the interquartile range (IQR) of the directivity can be used.
For two loudspeakers, this is done by calculating the localization curve twice using the weighted energy vector. First, one loudspeaker is weighted with the upper quartile of the directivity for its position and the other loudspeaker with the lower quartile. In the second calculation, this is done vice versa. The resulting area between both localization curves is interpreted as an estimate for the intersubjective variation.



(a) loudspeakers at $[0°, 45°]$ (b) loudspeakers at $[90°, 135°]$

Figure 27: Predicted localization curves (predicted angle vs. inter-channel level difference) for a 45° loudspeaker aperture angle at different positions using $\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}}$ with median (solid line with markers) and IQR (filled area) directivity weighting.

Figures 27(a) and 27(b) show localization curves and the estimated intersubjective

variation using the weighted energy vector with the broadband directivity, cf. Figure 17(b). They illustrate the larger variation of the localization for the lateral case. Obviously, the variation gets smaller for phantom source directions close to the loudspeakers. Similar tendencies can be found in the experimental results of Simon [SMR09]: Comparing the confidence intervals from his analysis with the variation predicted by the energy vector model yields a correlation of 0.87.

3.3.3 Phantom Source Localization in 3D

In the past years, the playback of phantom sources on three-dimensional loudspeaker arrangements has spread from research to practice, especially in cinemas. These arrangements are usually hemispherical [Ham05, TW11, NCW12]. However, there are only few studies about the perception of vertical phantom sources [FU99, Pul01a, KA12].

In contrast to localization of sound sources in the horizontal plane, localization in vertical planes uses monaural spectral cues [Bla83, Iid08] and works less accurately [Bla83]. There are models to predict the localization of real sound sources in vertical planes [Iid08, BMCQGAOB11, LB02]. Their applicability to vertical phantom sources is currently under investigation [BM13, WFZ13].

l	ϕ_l in °	θ_l in °	l	ϕ_l in °	θ_l in °
1	0	0.0	13	22.7	28.5
2	23.7	0.5	14	67.9	28.5
3	48.2	0.6	15	114.2	27.9
4	72.6	0.6	16	157.8	28.7
5	103.1	0.6	17	-156.4	28.7
6	138.5	0.2	18	-113.3	28.4
7	179.8	0.4	19	-65.4	28.5
8	-138.3	0.1	20	-22.7	28.0
9	-100.9	0.6	21	46.8	57.0
10	-69.8	0.4	22	133.4	57.0
11	-44.8	0.5	23	-133.4	56.6
12	-21.4	0.5	24	-43.4	57.7

Table 8: Azimuth angles ϕ_l and elevation angles θ_l of the 12+8+4 loudspeaker arrangement in the IEM CUBE.

Ambisonics with Height

In [BF11], we evaluated the localization of virtual sources and recordings using 1st and 4th order Ambi$^{\text{rE}}$, extending the studies in [BDG+07] to the third dimension. The excerpt presented here focuses on the virtual sources. The evaluation employed the hemispherical loudspeaker arrangement from Table 8 (an elevation angle of 0° refers to the horizontal plane) and the so-called AllRAD decoder design [ZF12] that is optimized for irregular arrangements. This arrangement is

permanently installed in the IEM CUBE and uses 24 Tannoy System 1200 loudspeakers, cf. background in Figure 12. It has a radius of about 5m and the lowest loudspeaker ring is at ear height.
The study evaluated multiple panning angles distributed all over the hemisphere using the pointing method from Section 2.3. The subjects' heads were not fixed, but they were instructed to face the 0° direction in the horizontal plane during the stimulus playback. The stimuli were pink noise bursts as in the previous experiments. Ten subjects participated in the listening experiment (not all of them were ELP members) and there were no repetitions.
Figure 28 shows the elevation angle error for 1st and 4th order Ambi$^{\text{rE}}$ merged into 3 groups of directions. As the number of directions is different in each group, the errorbars indicate interquartile ranges instead of confidence intervals.

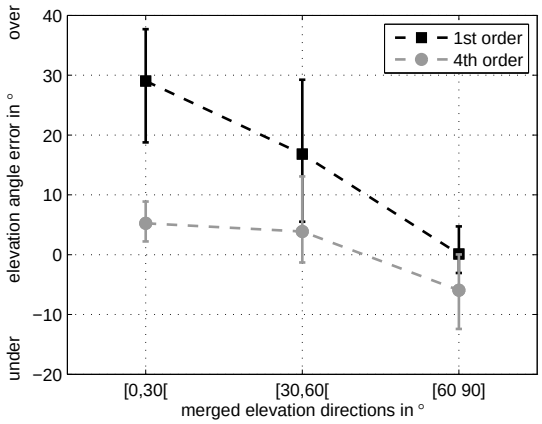


Figure 28: Elevation angle error for 1st and 4th order Ambi$^{\text{rE}}$. Directions are merged into 3 groups. Markers indicate medians and errorbars the interquartile ranges. Figure adapted from [BF11].

As expected, the errors are smaller for the 4th order reproduction. The overestimation of the height for directions near the horizontal plane can be explained by the hemispherical loudspeaker arrangement. The energy of the lower hemisphere is mostly preserved by the AllRAD decoder, but it is shifted to the loudspeakers in the horizontal plane. This results in active loudspeakers only in the horizontal plane and above. The upward shift is stronger for 1st order Ambisonics. The elevation angle of real sources near the north pole is typically underestimated [Bla83]. This holds true for the 4th order phantom sources. Using 1st order, the underestimation of elevation near the north pole is apparently compensated by the strong upward shift due to the loudspeaker arrangement.
The localization accuracy of noise using 4th order Ambi$^{\text{rE}}$ is even slightly better than the accuracy of speech using real sources [Bla83]. Therefore, it can be concluded that 4th order Ambi$^{\text{rE}}$ is sufficient for an accurate reproduction of elevation,

at least at the central listening position.

Pairwise Panning in the Median Plane

In order to study the vertical localization of phantom sources in a greater detail, a further experiment [WFZ13] evaluated the localization between two loudspeakers in the median plane. The experimental setup (loudspeaker type, distance to the listening position, reverberation time, stimulus, pointing method) was same as used in Section 3.2. The loudspeaker pair was placed at elevation angles of $\pm 20°$ behind an acoustically transparent screen. The different amplitude-panned phantom sources were created with 7 different inter-channel level differences (ICLD). Each of the 7 conditions was presented twice in random order.

The listening experiment was carried out with 15 subjects (not all of them were ELP members). Figure 29 presents the medians and confidence intervals of the perceived elevation angles for all tested ICLD values. There is a upward offset for the single loudspeakers ($|\text{ICLD}| = \infty\text{dB}$) and a saturation of the localization curve towards the loudspeakers. Small ICLD values exhibit a large subjective variation in the perceived angle. However, an ANOVA showed that except for the neighboring conditions (-6dB, -3dB) and (6dB, ∞dB), all differences are at least weakly significant ($p \leq 0.085$). This result agrees with the findings in [Pul01a, KA12] that vertical amplitude panning is possible but with a larger subjective variation compared to horizontal panning.

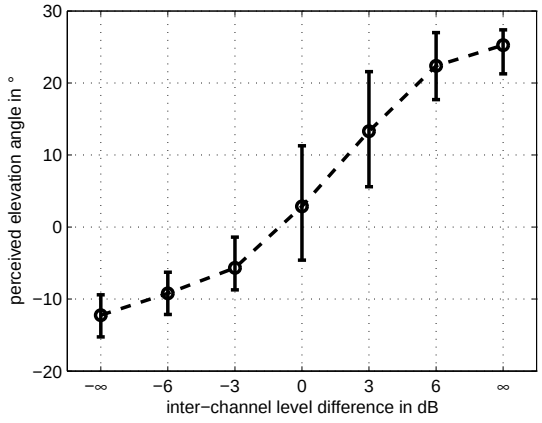


Figure 29: Median and corresponding 95% confidence intervals of the perceived elevation angle for different inter-channel level differences between a loudspeaker pair at $\pm 20°$ elevation angle in the median plane. Figure adapted from [WFZ13].

3.4 Conclusion on Phantom Source Localization

This chapter investigated the localization of phantom sources created by amplitude panning. First, it reviewed panning laws and experimental results for pairwise panning. The best predictions for broadband stimuli were achieved by the direction of the energy vector $\boldsymbol{r}_\mathrm{E}$. By incorporating the directivity of human hearing as direction-dependent weighting in the weighted energy vector $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$, the prediction was also possible for loudspeaker pairs that differ from the standard $\pm 30°$ arrangement.

Since localization experiments using amplitude panning on more than two loudspeakers are rare in the literature, own experiments were conducted using VBAP, MDAP, Ambi$^{\mathrm{r_E}}$, and Ambi$^{\mathrm{r_V}}$. All methods were evaluated at the central listening position using a regular arrangement of 8 loudspeakers. As VBAP always uses the smallest possible number of active loudspeakers, it yielded the smallest subjective variation in the perceived direction. However, the median localized directions of Ambi$^{\mathrm{r_E}}$ were closer to the ideal localization curve (perceived angle = panning angle). All predictors yielded practicable results independent of the number of active loudspeakers or the panning method. While the binaural model was the worst predictor, the best predictions were achieved with the intensity vector. The performance of the latter is only slightly better compared to $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$.

At the central listening position, there were no large differences between the tested panning methods. However, this does not hold for off-center listening position. Caused by the precedence effect, the localization is shifted towards the closest loudspeaker. The strong side lobes of Ambi$^{\mathrm{r_V}}$ even caused splitting of the phantom source into two events. Thus, it is important for good localization to keep the loudspeaker gains focused towards the target direction. In order to predict the source splitting, the models would have to use probabilistic assessment instead of estimating a single output angle.

The chapter also investigated the cause of the subjective variation in phantom source localization and presented a simple method to estimate this intersubjective variation by using the variation in the directivity of human hearing.

The studies presented about three-dimensional amplitude panning showed that panning in vertical directions is also possible. However, the localization accuracy is worse than in the horizontal plane which is also the case for real sound sources.

Chapter IV

PHANTOM SOURCE WIDTH

In a typical concert hall, early reflections from the walls and the ceiling influence the perceived spatial extent of a sound source, e.g. an instrument or a voice. This perceived size of a sound source is often called *auditory* or *apparent source width* (ASW). Some technical measures are known to correlate well with the ASW in rooms [SGS74, BM81, BL86a, HBO95, MI05]. The two most common measures are the inter-aural cross correlation coefficient (IACC) and the lateral energy fraction (LF), calculated from recordings using a dummy head and combination of an omni-directional and a figure-of-eight microphone, respectively.
For single loudspeakers and multiple loudspeakers that play decorrelated signals in an anechoic chamber, the amount of low frequencies and the overall level increase the perceived source width [CT03]. Differences in perceived source width are more distinct for broadband stimuli [SP11]. Furthermore, an increase of the stimulus duration increases the source width [PB82, HP06]. In [Hir08], a time constant for the perception of source width in the range of 40-80ms was suggested which corresponds to the integration times for IACC and LF. Correlated signals, as in amplitude panning, yield narrower sources [ZFMS11].
There are only few studies about phantom source width using amplitude panning [Pul99, MWCQ99b] and the question of how the number of active loudspeakers, their gains, and their spatial distribution influence phantom source width has not been addressed systematically, yet. This chapter presents studies about these influences while keeping duration, loudness, and frequency content of the stimuli constant.
Section 4.1 starts the studies by investigating source width using differently wide arrangements of two and three loudspeakers. The perceived width is evaluated in a listening experiment. Furthermore, the section compares the experimental results to known measures from room acoustics and introduces alternative predictors. Section 4.2 extends the previous generic study and its predictors to typical amplitude-panning scenarios in the horizontal plane using VBAP, MDAP, and Ambisonics. Finally, Section 4.3 presents a brief overview about further research in phantom source width.

4.1 Source Width using 1, 2, 3 Loudspeakers

Studies show that the loudspeaker aperture angle has an influence on the perceived source width [FMS11, KP11]. To improve the understanding of this relationship, this section presents a study with differently wide loudspeaker arrangements in a listening setup with dominant direct sound. The loudspeaker arrangements are symmetrical to the 0° axis and use one, two, or three loudspeakers that play the same signal at the same time.

The first part of this section describes a listening experiment that evaluates the perceived source width. The impulse responses of the loudspeakers in the experiment setup were measured with a dummy head and a microphone array. From these measurements, IACC and LF measures are derived and compared to the results of the listening experiment. The section also presents a modification of LF to make it a valid predictor for the experimental data. Furthermore, it introduces a simple model for the perceived source width that predicts the data without acoustic measurement, just based on the gains and directions of the loudspeakers.

4.1.1 Frontal Phantom Source Width for Symmetrical Loudspeaker Arrangements (Exp. 1)

The evaluation studies the perception of source width for one, two, and three active loudspeakers playing the same signal and having various aperture angles between them. Figure 30 shows the experimental setup with 17 loudspeakers: loudspeaker 0 at 0°, and loudspeaker pairs $1 \ldots 8$ at $\pm 5° \ldots \pm 40°$. Except for the loudspeaker positions, the whole experimental setup is identical to the setup presented in Section 2.2. All measurements in the remainder of this section were using the same setup and conditions.

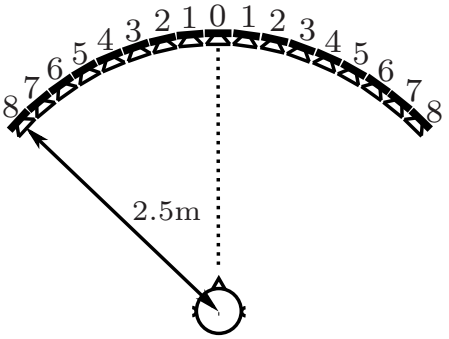


Figure 30: Loudspeaker setup used in experiment 1 [Fra13].

loudspeaker(s)	0	1	2	3	4	5	6	7	8
angle(s) ϕ_l in °	0	±5	±10	±15	±20	±25	±30	±35	±40
0	1								
10		$1/\sqrt{2}$							
20			$1/\sqrt{2}$						
30				$1/\sqrt{2}$					
40					$1/\sqrt{2}$				
50						$1/\sqrt{2}$			
60							$1/\sqrt{2}$		
70								$1/\sqrt{2}$	
80									$1/\sqrt{2}$
$C10$	$1/\sqrt{3}$	$1/\sqrt{3}$							
$C20$	$1/\sqrt{3}$		$1/\sqrt{3}$						
$C30$	$1/\sqrt{3}$			$1/\sqrt{3}$					
$C40$	$1/\sqrt{3}$				$1/\sqrt{3}$				
$C50$	$1/\sqrt{3}$					$1/\sqrt{3}$			
$C70$	$1/\sqrt{3}$							$1/\sqrt{3}$	
$C80$	$1/\sqrt{3}$								$1/\sqrt{3}$

Table 9: Conditions in the experiment 1: loudspeaker indices, directions ϕ_l, and gains g_l (gains of zero are not shown).

Table 9 shows the angles and gains of the loudspeakers for each of the 16 conditions. Empty entries in the table mean that the corresponding loudspeakers are not active. Condition 0 is a single loudspeaker playing from 0° in front of the listener. Conditions $10 \ldots 80$ correspond to 2-channel stereophony using frontally centered pairs of loudspeakers with the same amplitude gain and aperture angles ranging from 10° to 80°. In conditions $C10 \ldots C80$ the center loudspeaker (loudspeaker 0 at 0°) is added with the same gain. These additional conditions aim at extending the applicability of the relationships obtained to arbitrary amplitude panning methods that use more than two loudspeakers, such as MDAP or Ambisonics. Condition $C60$ has not been tested due to an error in the playback software. The gains in all conditions were normalized to a constant overall energy which results in gains that depend on the number of active loudspeakers. Furthermore, the symmetrical arrangement aims at creating a phantom source direction of 0° for all conditions in order to exclude differences in the localization direction between the conditions.

Fourteen subjects participated in the listening experiment. Each of the 16 conditions was presented seven times for each subject in random order. The stimulus was 1.5s of pink noise at a level of 65dB(A). The subjects were allowed to repeat the stimulus at will by pressing a button on a keyboard. They were asked to measure the perceived source width in terms of an index and to write their answer on a questionnaire. The index expressed the perceived width in terms of numbers on the increasingly wide, nested loudspeaker pairs according to Figure 30. It was

also possible to use half indices to rate perceived widths that are between adjacent indices which results in a possible resolution of 5°. The subjects were told to face forward but there was no head fixation.

Results

The answers were averaged over all subjects and all repetitions. Figure 31 shows the resulting mean value and corresponding 95% confidence interval of the perceived source width for each condition. The results in the figure were arranged to ascending aperture angles between the two outmost active loudspeakers.

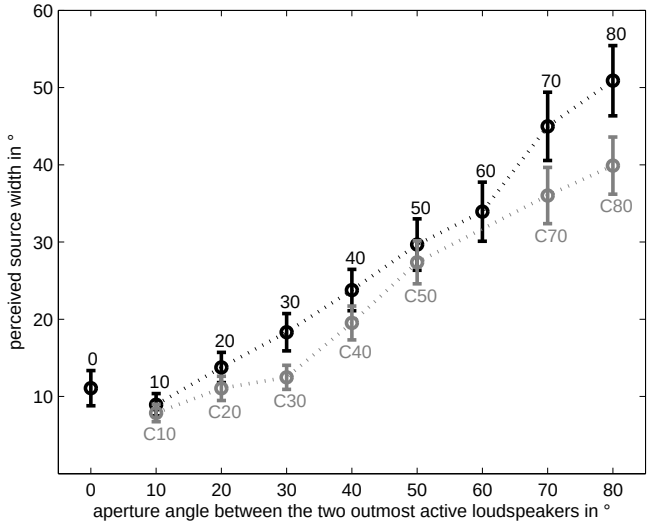


Figure 31: Mean and 95% confidence interval of the perceived source width for each condition, arranged to ascending aperture angle between the two outmost active loudspeakers [Fra13].

Within the 2-loudspeaker conditions $10 \ldots 80$, an increase of the aperture angle yields an increase of the perceived source width. An ANOVA confirms the aperture angle as a highly significant factor ($p \ll 0.001$). This holds true for the conditions with the additional center loudspeaker $C10 \ldots C80$. The phantom source width is smaller than the physical width enclosing all active loudspeakers, which agrees with the findings in [SP11] using decorrelated loudspeaker signals. Comparing both groups of results, these with and those without the center loudspeaker, the addition of the center loudspeaker yields a highly significant decrease of the perceived source width. This relation agrees with the findings in [KP11], and it can be explained as the active center loudspeaker decreases the relative share of lateral sound. Direct comparisons of the corresponding condition pairs (with and without center loudspeakers) showed that differences between the mean values of $C10/10$ and $C50/50$ are insignificant ($p > 0.05$).

Interestingly, conditions $C10$, 10, and $C20$ yield smaller mean values for the perceived source width than the single center loudspeaker 0. However, the mean values are not significantly different. Hence, the lower bound for perception of source width is about 10° in our experimental setup. This bound is expected to be dependent on the acoustical properties of the room and may differ for other rooms.

4.1.2 Technical Measures and Predictors for Source Width

Inter-Aural Cross Correlation Coefficient (IACC)

In order to calculate the inter-aural cross correlation coefficient (IACC) for the tested conditions, binaural impulse responses were measured for each loudspeaker using a dummy head. For each condition, h_{left} and h_{right} are the impulse responses of the left and the right ear of the dummy head, respectively. They are calculated as the linear superposition of the binaural impulse responses for each loudspeaker $h_{l,\text{left}}$ and $h_{l,\text{right}}$ with the appropriate loudspeaker gains g_l according to Table 9

$$h_{\text{left}} = \sum_{l=1}^{\text{L}} h_{l,\text{left}}\, g_l \qquad \text{and} \qquad h_{\text{right}} = \sum_{l=1}^{\text{L}} h_{l,\text{right}}\, g_l. \tag{14}$$

Both impulse responses are convolved with A-weighted pink noise resulting in the ear signal s_{left} and s_{right}. The IACC is defined as the maximum of the inter-aural cross correlation function (IACF), cf. [ISO09]

$$\text{IACF}(\tau) = \frac{\int_{t_1}^{t_2} s_{\text{left}}(t) s_{\text{right}}(t+\tau)\text{d}t}{\sqrt{\left[\int_{t_1}^{t_2} s_{\text{left}}^2(t)\text{d}t\right]\ \left[\int_{t_1}^{t_2} s_{\text{right}}^2(t)\text{d}t\right]}}, \tag{15}$$

$$\text{IACC} = max_{\tau\in[-1\text{ms};1\text{ms}]}|\text{IACF}(\tau)|, \tag{16}$$

Typically, the observation time is set to the first 80ms, i.e. $t_1 = 0$ms and $t_2 = 80$ms. As this version of the IACC considers only the early part of the impulse responses, it is called early IACC or IACC_{E}. Furthermore, the IACC is commonly not calculated for the broadband signals, but separately for three octave bands around 500Hz, 1kHz, and 2kHz [HBO95], instead. The three correlation coefficients are averaged. Here the early IACC is employed for the three octave bands, denoted as IACC_{E3}.

There are different values for the perceptually just noticeable difference (JND) of IACC_{E3} in the literature: 0.075 [ISO09, CDL93], 0.05–0.08 [Oka02], 0.038 [Bla02]. Figure 32 shows the computed $1 - \text{IACC}_{\text{E3}}$ measures of all conditions in relation to the mean of the perceived source width from the listening experiment. The IACC_{E3} values range from 0.82 to 0.97 and cover 2–4 JNDs which predicts a poor

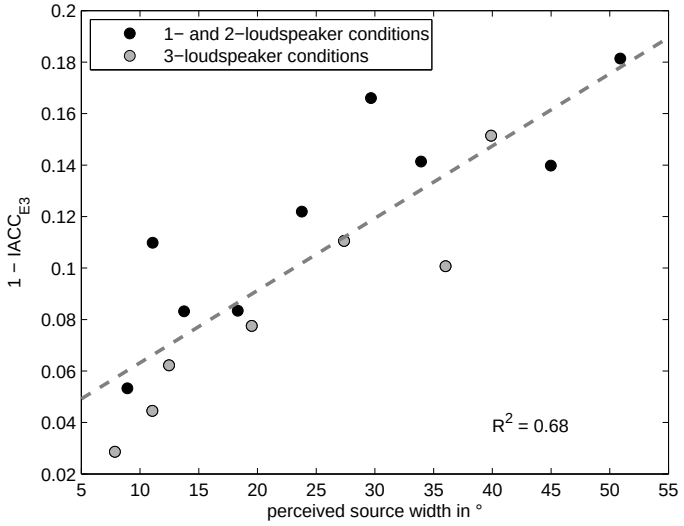


Figure 32: Regression of $1 - \mathrm{IACC_{E3}}$ to the listening experiment results. Figure adapted from [Fra13].

discriminability, what contradicts the results of the listening experiment, cf. Figure 31. The value of $\mathrm{R}^2 = 0.68$ for the coefficient of determination indicates only a fair correlation between the experimental results and the objective predictor. Interestingly, the prediction of the 3-loudspeaker conditions is better than of the 1- and 2-loudspeaker conditions. Altogether, $\mathrm{IACC_{E3}}$ does not seem to be an optimal predictor of the perceived source width in case of simultaneous sound incidence. In other cases, the $\mathrm{IACC_{E3}}$ is a better predictor if the temporal structure of the loudspeaker signals is manipulated, e.g. by decorrelation algorithms [ZFMS11], cf. Section 4.3.

Lateral Energy Fraction (LF)

The lateral energy fraction (LF) is also used to describe width [BM81, HBO95]. It is derived from the impulse response measurements using an omni-directional microphone and a figure-of-eight microphone with the receiving direction to the side. This yields the impulse responses $h_\circ$ and h_∞, respectively. For each condition, both responses are computed from the linear superposition of the individual impulse responses of each loudspeaker $h_{l,\circ}$ and $h_{l,\infty}$ with the appropriate loudspeaker gains g_l according to Table 9

$$h_\circ = \sum_{l=1}^{\mathrm{L}} h_{l,\circ}\, g_l \qquad \text{and} \qquad h_\infty = \sum_{l=1}^{\mathrm{L}} h_{l,\infty}\, g_l. \tag{17}$$

Just as in the computation of the IACC, the impulse responses are convolved with A-weighted pink noise yielding the signals s_∞ and $s_\circ$. The lateral energy fraction

is defined by:

$$\mathrm{LF} = \frac{\int_{t_0}^{80\mathrm{ms}} s_\infty^2 \mathrm{d}t}{\int_{0\mathrm{ms}}^{80\mathrm{ms}} s_\circ^2 \mathrm{d}t}. \tag{18}$$

As the upper integration bound is normally set to 80ms, the measure is sometimes also called early lateral energy fraction [CDL93]. According to ISO3382 [ISO09], the lower integration bound of the figure-of-eight signal is defined as $t_0 = 5$ms. Although some authors calculate the LF in three octave bands around 500Hz, 1kHz, and 2kHz, i.e., similar as with the $\mathrm{IACC}_{\mathrm{E3}}$, the broadband version of the LF is used in this work.
Literature gives the following values for the JND of the LF: 0.048 (computed) and 0.058 (measured)[CDL93], 0.075 [ISO09], and 0.045–0.07 [Bla02].

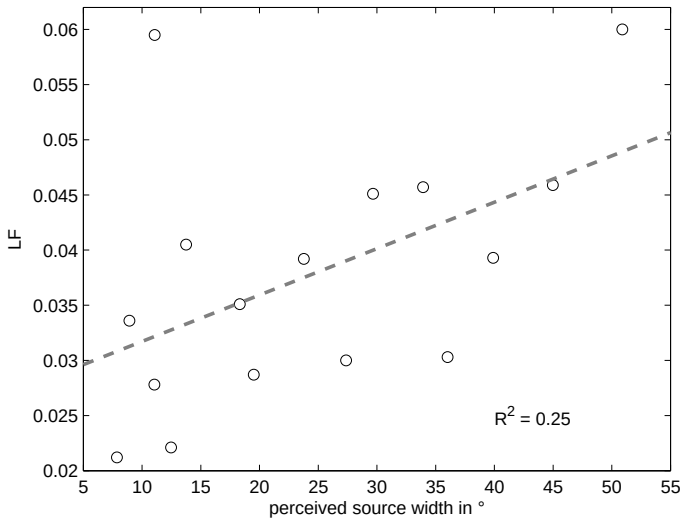


Figure 33: Regression of the standard LF measures ($t_0 = 5$ms) to the listening experiment results. Figure adapted from [Fra13].

Figure 33 shows that the standard LF measure is not related ($\mathrm{R}^2 = 0.25$) to the experimental results. What is more, the LF values range from 0.02–0.06 and lie within one JND only, what contradicts the experimental results. The lower integration bound of $t_0 = 5$ms for the figure-of-eight signal yields an exclusion of the direct part of the sound. Thus, these LF values represent the effect of the early reflections exclusively that is perceptually independent of the condition.

In order to improve the suitability of the LF for direct sound, the lower integration bound t_0 of the figure-of-eight signal s_∞ is changed to $t_0 = 0$ms. However, the value of $\mathrm{R}^2 = 0.27$ brings virtually no improvement.
The applicability is still prevented by signal cancellation of the simultaneous sound

incidence at the figure-of-eight microphone. The next paragraph presents an approach to overcome the poor prediction by a further modification of the LF measurement.

Energetic superposition

The signal cancellation is due to the symmetrical loudspeaker arrangement and the single measurement position. A listener would not perceive this cancellation, as humans have two ears and the head as a diffraction body in between. In order to avoid destructive interference also for the LF measure, the impulse responses of the loudspeakers are superimposed energetically

$$h_{\circ} = \sqrt{\sum_{l=1}^{\mathrm{L}} (h_{l,\circ}\, g_l)^2} \qquad \text{and} \qquad h_{\infty} = \sqrt{\sum_{l=1}^{\mathrm{L}} (h_{l,\infty}\, g_l)^2}. \tag{19}$$

This was done for all conditions and the LF measures were calculated again. Figure 34 shows that the correlation between these values and the perceived source width is remarkably high ($\mathrm{R}^2 = 0.96$). The LF measure now ranges from 0.06–0.55 and covers 7–10 JNDs, which is comparable to the significant differences in the listening experiment results.

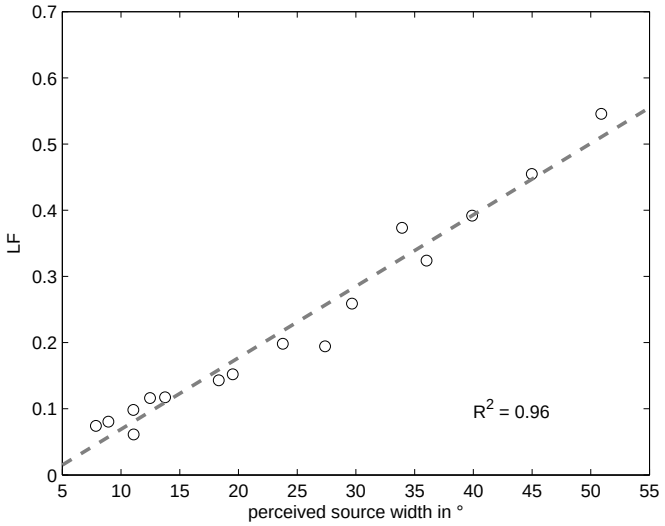


Figure 34: Regression of the energetically superimposed LF measures ($t_0 = 0$ms) to the listening experiment results. Figure adapted from [Fra13].

Nevertheless, the energetically superimposed lateral energy fraction cannot be measured by a single measurement of simultaneously active loudspeakers as the superposition in the sound field is linear. In order to avoid signal cancellation for the case of simultaneously active loudspeakers, an alternative definition of the LF has to be found.

Multiple measurement positions

A pair of equally loud coincident signals from the symmetric pair of loudspeakers cancels in the figure-of-eight microphone due to its pickup pattern. The cancellation is not entirely destructive when measuring at an off-center position. This is because the pair of loudspeaker signals arrives with unequal time-delays, i.e., with a phase difference that linearly grows with frequency. Thus cancellation can be reduced by measuring at N_{pos} positions on a line that is shifted along the axis of the figure-of-eight microphone and averaging the LF_n values hereby obtained

$$\mathrm{LF} = \frac{1}{N_{pos}} \sum_{n=1}^{N_{pos}} \mathrm{LF}_n. \tag{20}$$

Each of the LF_n values results from a linear superposition of the loudspeaker signals as in Eq. (17).

In a diffuse sound field, the correlation between two positions with a distance of d is given by $\frac{\sin(kd)}{kd}$. The wavenumber k is defined as $k = {}^{2\pi f}/_{\mathrm{c}}$, with $\mathrm{c} = 343\ {}^{\mathrm{m}}/_{\mathrm{s}}$. The first zero of the correlation function is at $f = \frac{\mathrm{c}}{2d}$. At this frequency, sound at the two positions is completely decorrelated. Although the examined sound field is not diffuse in the experimental condition, the above-mentioned relation is applied as an estimate for the frequency f_{min} above which the displacement of the measurement positions is effective. The frequency depends on the measurement aperture d_{max}, i.e. the distance between the outmost measurement positions as

$$f_{min} \approx \frac{\mathrm{c}}{2d_{max}}. \qquad \text{with } \mathrm{c} = 343\ {}^{\mathrm{m}}/_{\mathrm{s}} \text{ in air.} \tag{21}$$

Figure 35 shows the measurement setup with a figure-of-eight microphone and a omni-directional microphone at each measurement position.

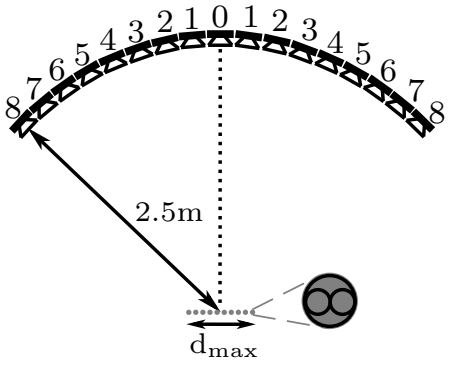


Figure 35: Measurement setup with multiple measurement positions with an aperture of d_{max}. Each measurement position has an omni-directional and a figure-of-eight microphone.

A simulation was done to evaluate how good the energetic superposition is approximated by averaging over multiple measurement positions where loudspeaker

signals are superimpose linearly. These measurement positions were simulated by appropriate delaying (rounded to integer samples at a sampling rate of 44.1kHz) and level adjustment of the measured impulse responses from the central listening/measurement position. As before, the impulse responses were convolved with A-weighted pink noise. The simulation ignored the influence of the loudspeaker directivity which is assumed to be negligible for the used range of d_{max}. In the simulation, the number of measurement positions N_{pos} and the size of the measurement aperture d_{max} was varied. The maximum deviation between the LF measures of the energetic superposition and the average linear superposition at multiple measurement positions was used as quality measure for the approximation. To ensure that there is no perceptible difference in the approximation, the deviation must be $\leq 1/2$ JND of the LF which is approximately 0.03. In Figure 36, the absolute value of the deviation is presented in gray scale. Whenever the deviation is below the value of 0.03, the area is white. Darker areas represent larger deviations.

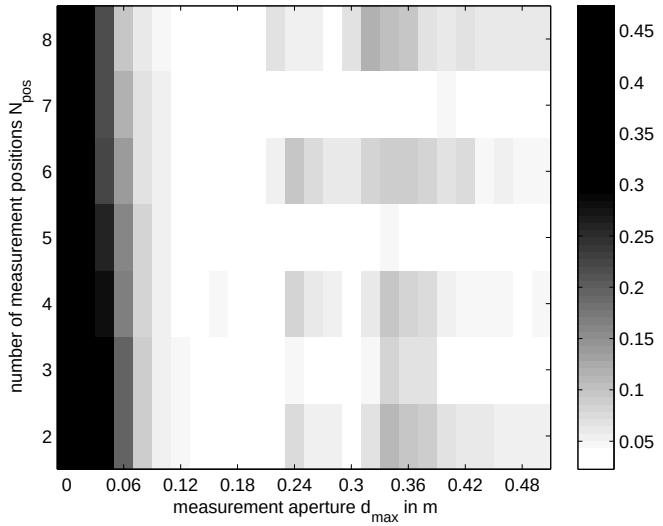


Figure 36: Maximum deviation of the approximation of the energetically superposition by averaging over multiple linear superimposed measurement positions in dependence of the number of measurement positions N_{pos} and the aperture d_{max}. Figure adapted from [Fra13].

The number of measurement positions N_{pos} has a weak influence on the maximum deviation. For apertures $d_{max} \geq 0.14$m, the deviations are smaller than 0.03, even when averaging over only $N_{pos} = 2$ positions. This distance is similar to the head diameter and results in an lower frequency bound for the avoidance of destructive interference of $f_{min} = 1.2$kHz. For phantom source width, higher frequency components seem to be important. This result agrees to the findings in [BL86b, MM88] about higher frequency components also contributing to the

perception of source width in concert halls.
The regression of the experimental results to the averaged LF measures over $N_{pos} = 2$ positions, $d_{max} = 0.14$m (± 0.07m apart from the center of the arrangement), yields an R^2 of 0.95, which is similar to the energetically superimposed LF measures, cf. Figure 34.
The advantage of multiple measurement positions can also be found in other fields of acoustics. In recording technique, spaced main microphone arrays are preferred over coincident arrays when capturing spatial impressions [The91]. The literature about room acoustics reports large spatial fluctuation of room acoustic measures in concert halls that occur only in measurements, but not in perception [dVHB01, vDS11].
In the remainder of this work, the standard LF measure (1 position, linear superposition, $t_0 = 5$ms) is labeled with $\mathrm{LF}^{\mathrm{ISO}}$ and the energetically superimposed LF with LF^2.

Energy Vector ($\boldsymbol{r}_\mathrm{E}$)
The direction of the energy vector $\boldsymbol{r}_\mathrm{E}$ [Ger92], cf. Eq. (9), was used as a predictor of the perceived direction of a phantom source in Chapter 3. Here, its magnitude is proposed as predictor of the perceived source width. A magnitude value of 1 indicates that only one loudspeaker is active, while one of 0 corresponds to energy distributed in all directions or in opposing directions. In the tested conditions, all L active loudspeakers are driven by $g_l = 1/\sqrt{\mathrm{L}}$ so that the overall energy is normalized $\sum_{l=1}^{\mathrm{L}} g_l^2 = 1$. In this case, $\boldsymbol{r}_\mathrm{E}$ is the average of the direction vectors $1/\mathrm{L} \sum_{l=1}^{\mathrm{L}} \boldsymbol{\theta_l}$. For the tested, frontal conditions, the magnitude of the energy vector is strongly related to the lateral energy fraction under free-field conditions. Therefore, the correlation of the magnitude of the energy vector to the perceived width of phantom sources reported in [FMS11] is not surprising.

Figure 37 underlines the excellent correlation ($R^2 = 0.97$) between the experimental results and the magnitude of the energy vector. The regression coefficients yield a formula for the prediction of the perceived source width

$$\alpha = 186.4^\circ \cdot (1 - ||\boldsymbol{r}_\mathrm{E}||) + 10.7^\circ. \tag{22}$$

The additive bias of 10.7° relates to a lower bound of the perceived source width that was found in the listening experiment, cf. Figure 31. Moreover, it is related to the bias of the LF measure that is caused by early reflections and therefore independent of the condition, cf. Figure 33. Naturally, the exact value of this bias is depending on the room. Nevertheless, the linear relation between the length of

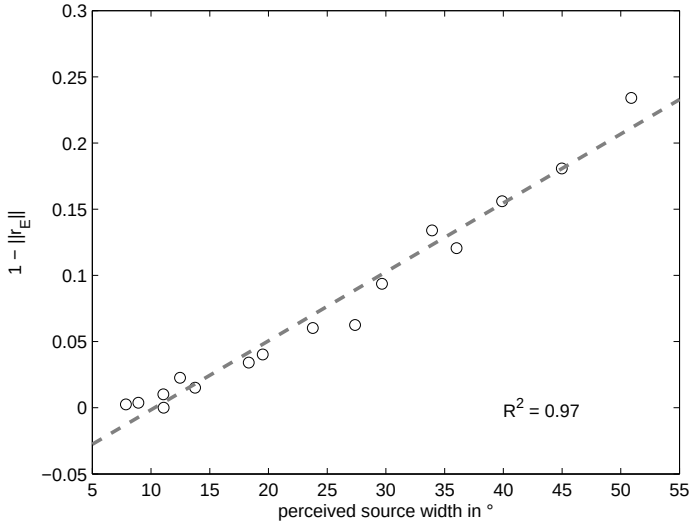


Figure 37: Regression of $1 - ||\boldsymbol{r}_{\mathrm{E}}||$ to the listening experiment results [Fra13].

the energy vector and the perceived source width is valid for other rooms, as long as the listener sits centrally and in the direct sound field of the loudspeakers.
Note that the length of the weighted energy vector that is introduced as a predictor for source width in Section 4.2, yields a sightly improved R^2 value of 0.98.

4.1.3 Discussion

This section studied the source width of frontal phantom sources created by one, two, and three loudspeakers playing the same signal in a listening setup with dominant direct sound. For the two-channel conditions, a listening experiment revealed a monotonic relation between the physical width of the active loudspeaker pair and the perceived width of the phantom source. This relation was also demonstrated to work in case of an additional center loudspeaker. Comparing both groups of conditions, the addition of the center loudspeaker decreases the source width. This finding agrees with results from the literature [KP11].
The experimental results were compared to the early inter-aural cross correlation coefficient ($\mathrm{IACC}_{\mathrm{E3}}$) and the lateral energy fraction (LF). The $\mathrm{IACC}_{\mathrm{E3}}$ yields a fair correlation to the listening experiment results, whereas the correlation to the standard $\mathrm{LF}^{\mathrm{ISO}}$ is poor. That is because the $\mathrm{LF}^{\mathrm{ISO}}$ measure excludes the direct sound and solely considers the early reflections whose amount was found to be condition-independent in our listening setup. Modifications were presented that improve the LF measurement for simultaneous incidence of direct sound and symmetrical loudspeaker arrangements. This was achieved by reducing the lower bound for the integration of the figure-of-eight microphone and avoiding signal cancellation at the figure-of-eight microphone. The optimal solution would be

an energetic superposition of the loudspeaker signals (LF^2) at the microphone which is impossible in practice when multiple loudspeakers are playing simultaneously. Enabling such conditions, a more versatile measurement approximates this by recording the LF at multiple positions. Averaging of two positions at a distance of ±7cm is sufficient. As this distance avoids signal cancellation only above 1.2kHz, the importance of high frequency components for source width is obvious. Despite the improved prediction of source width under listening conditions with dominant direct sound, the adapted LF measures are still applicable for measurements in reverberant rooms. In this work, $IACC_{E3}$ and LF measures were computed from impulse responses convolved with A-weighted pink noise. Note that similar results were achieved when applying solely the impulse responses as in [Fra13].

Finally, the simplest predictor of the perceived source width is solely based on loudspeaker directions and gains. This energy vector model assumes typical studio conditions where the listener sits centrally within the effective critical distance, i.e., the direct sound is more prominent than the reflections. The condition-independent lower bound of source width is caused by reflections and maps to an additive constant in the model. The exact value of this bias is depending on the listening environment.

4.2 Source Width using Amplitude Panning with Multiple Loudspeakers

This section extends the previous study to amplitude-panning methods by four listening experiments. The experiments evaluate the perceived source width of phantom sources for panning angles that lie between and on the loudspeakers. Although different panning angles relative to the loudspeakers are evaluated, the loudspeakers are arranged in a way that provides a phantom source localization at (experiments 2 and 3) or close to (experiments 3 and 4) the same direction, independent of the panning angle. This section also shows regressions of the experimental results to technical measures and predictors.

Setup and Method

Figure 38 shows the experimental setup with 21 loudspeakers that was used for experiments 2-5. The stimulus was 1.5s of pink noise presented at 65dB(A). A full pairwise comparison of the conditions was done in all experiments. After a short training phase, all comparisons were repeated four times. The subjects responded by pressing the button on a keyboard that selects the sounds in the pair that appeared to be wider. In experiments 2 and 3, the subjects could listen to each sound in the pair only once. A total of 32 subjects participated in the experiments, 16 subjects in each experiment. Each subject participated either in experiment 2 or 3 and in experiment 4 or 5.

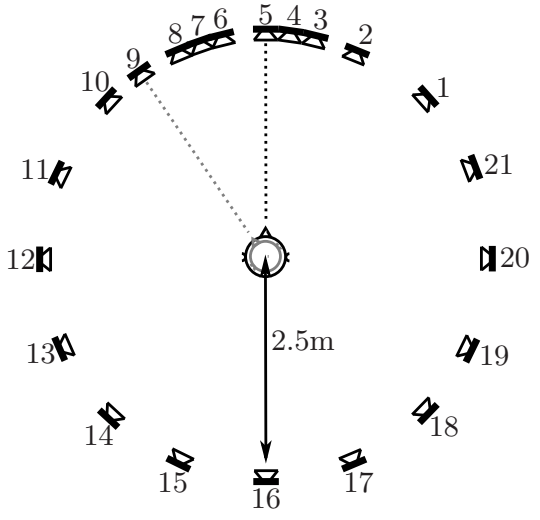


Figure 38: Experimental setup for experiments 2-5; 33° head orientation for experiment 3 is drawn in gray.

The experiments used different amplitude-panning methods: VBAP, MDAP, Ambi$^{\text{rv}}$, and Ambi$^{\text{re}}$. All methods were applied to regular equally spaced loudspeaker arrangements with two different numbers of loudspeakers: one arrangement with 8

loudspeakers and a loudspeaker aperture angle of $\Delta\phi_L = 45°$ and the other one with 16 loudspeakers and an aperture angle of $\Delta\phi_L = 22.5°$. Ambisonics was always used with the highest possible order, i.e. N = 3 for conditions with 8 loudspeakers and N = 7 for conditions with 16 loudspeakers. MDAP used B = 10 panning directions uniformly distributed within a spread of $\phi_{\mathrm{MDAP}} = {}^1\!/_2\Delta\phi_L$. On the regarded loudspeaker arrangements, the number of conditions could be reduced as some methods yield exactly the same loudspeaker gains for specific panning angles.

4.2.1 Frontal Phantom Source Width for Amplitude Panning (Exp. 2)

This experiment investigates the width of phantom sources from the front. The subjects were facing forward, cf. black head symbol in Figure 38. There was no head fixation in order to allow small unconscious head movements [Bla83] that are important for localization [Mac08] and spatial impression [BKM07].

Conditions

The experiment evaluates the source width of VBAP, MDAP, Ambi$^{\mathrm{rE}}$, and Ambi$^{\mathrm{rV}}$ on regular arrangements of 8 and 16 loudspeakers, resulting in loudspeaker aperture angles of $\Delta\phi_L = 22.5°$ and 45°, respectively. The phantom sources are either panned on the angle of the frontal loudspeaker (conditions 1, 2, 3, 4, 6, 8, and 9 in Table 10) or half way between the frontal loudspeaker pairs (conditions 5 and 7). This provides a localization at the 0° direction for all conditions independent of the panning angle.
In order to reduce the number of conditions and pairwise comparisons, Ambi$^{\mathrm{rV}}$ is only evaluated for panning on the frontal loudspeaker. Further reduction is possible as conditions 5 and 7 represent VBAP, MDAP, and Ambi$^{\mathrm{rE}}$ panned half way between the loudspeakers for $\Delta\phi_L = 22.5°$ and 45°, respectively. As reference, a real sound source, i.e. a single loudspeaker, is employed in condition 1. This condition also represents a VBAP source panned on this loudspeaker. All conditions that represent multiple panning methods are marked by asterisks.

Results

Within-subject repetitions were averaged and the pairwise comparison matrices were transformed into scale values [Thu94], yielding 16 individual scales. Additionally, a scale is calculated from the pooled comparison matrix from all subjects. Figure 39 presents the median values and the corresponding confidence intervals of the individual scales and the pooled scale. The median and the pooled scale correlate with 0.98.

condition		* 1	2	3	4	* 5	6	* 7	8	9
aperture $\Delta\phi_L$ in °		-	22.5	22.5	22.5	22.5	45.0	45.0	45.0	45.0
angle rel. $\Delta\phi_L$		0	0	0	0	1/2	0	1/2	0	0
panning method		real	MDAP	Ambi$^{\mathrm{rE}}$	Ambi$^{\mathrm{rV}}$	VBAP	Ambi$^{\mathrm{rV}}$	VBAP	MDAP	Ambi$^{\mathrm{rE}}$
number	**ϕ_L in °**	**loudspeaker gains for each condition in experiment 2 and 3**								
1	-45.0			-0.06	-0.06		0.13		0.19	0.31
2	-22.5		0.19	0.30	0.06			0.71		
3	-11.3					0.71				
4	-5.6									
5	0.0	1.00	0.96	0.90	0.97		0.94		0.96	0.89
6	11.3					0.71				
7	16.8									
8	22.5		0.19	0.30	0.06			0.71		
9	33.0									
10	45.0			-0.06	-0.06		0.13		0.19	0.31
11	67.5			0.03	0.06					
12	90.0			-0.02	-0.06		-0.13			-0.07
13	112.5			0.01	0.06					
14	135.0			-0.01	-0.06		0.13			0.04
15	157.5			0.01	0.06					
16	180.0			-0.01	-0.06		-0.13			-0.04
17	-157.5			0.01	0.06					
18	-135.0			-0.01	-0.06		0.13			0.04
19	-112.5			0.01	0.06					
20	-90.0			-0.02	-0.06		-0.13			-0.07
21	-67.5			0.03	0.06					

Table 10: Loudspeaker directions and gains for each condition of experiment 2 (frontal phantom source width for amplitude panning) and 3 (lateral phantom source width for amplitude panning). Conditions 1, 5, 7 represent multiple panning methods. All gains $|g_l| < 0.005$ are not shown.

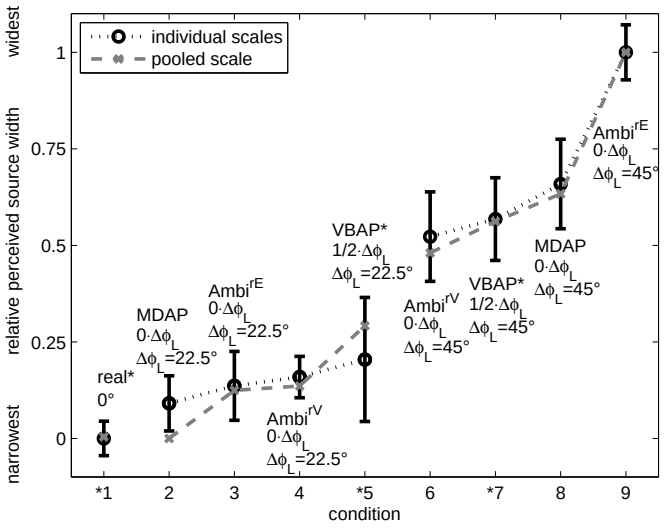


Figure 39: Results of experiment 2 (frontal phantom source width for amplitude panning): Median values and corresponding 95% confidence intervals of individual scales and pooled scale. Median scale and pooled scale are normalized between 0 and 1.

There is a significant effect of the two loudspeaker aperture angles $\Delta\phi_L$ on the perceived source width ($p < 0.001$). Condition 1 is perceived significantly narrower than all other conditions ($p < 0.05$), except for conditions 2 and 3. Within the conditions with $\Delta\phi_L = 22.5°$, the mean values of conditions 2/5 are significantly

different ($p = 0.03$), whereas the differences between conditions 2/4 ($p = 0.051$) and 4/5 ($p = 0.076$) are weakly significant. Condition 9 is perceived significantly wider than all other conditions ($p < 0.005$). Aside from that, there are no significant differences between the conditions with $\Delta\phi_L = 45°$, except for conditions 6/8 that differ weakly significantly ($p = 0.095$).

For VBAP, the perceived width for panning angles between the loudspeakers or aligned with one loudspeaker is significantly different (conditions 1/5 and 1/7). For Ambi$^{\mathrm{r_E}}$, the dependency of the perceived width on the panning angle is only noticeable in the case of $\Delta\phi_L = 45°$ (conditions 7/9), and for MDAP, it is only noticeable in the case of $\Delta\phi_L = 22.5°$ (conditions 2/5) and otherwise panning-independent. The difference between the two different Ambisonics weightings (conditions 6/9) for panning angles aligned with one loudspeaker is exclusively significant for the larger loudspeaker aperture angle $\Delta\phi_L = 45°$.

Table 11 presents the R^2 values for linear regressions of the technical measures and predictors to the pooled scale and the median of the individual scales of the experimental results. The $IACC_{E3}$ is the best predictor for experiment 2, followed by the energetically superimposed lateral energy fraction LF^2. The unweighted energy vector $\boldsymbol{r}_E$ seems to be a poor predictor.

$\mathbf{R^2}$	$IACC_{E3}$	LF^2	$\boldsymbol{r}_E$
pooled	0.82	0.64	0.30
median	0.85	0.64	0.32

Table 11: R^2 values for linear regression of technical measures against the results of experiment 2 (frontal phantom source width for amplitude panning).

The length of $\boldsymbol{r}_E^w$ as source width predictor

The regression results indicate that the energy vector $\boldsymbol{r}_E$ is not a suitable predictor for the width of all conditions. It overestimates the width of Ambi$^{\mathrm{r_V}}$, whereas it underestimates the width of MDAP. The conditions using Ambi$^{\mathrm{r_V}}$ employ the highest levels on the rear loudspeakers, especially in the case of the larger loudspeaker aperture angle $\Delta\phi_L = 45°$, cf. Table 10. The overestimation of the width could be reduced by reducing the influence of the rear loudspeakers when computing the energy vector. MDAP employs the center loudspeaker at 0° and two symmetrically arranged loudspeakers at more lateral positions. Increasing the influence of these lateral loudspeakers would reduce the underestimation of the width.

In order to increase the predictive accuracy of the energy vector model, a direction-dependent weighting is necessary that reduces the sensitivity for rear directions and increases the sensitivity for lateral directions. This necessity initiated the introduction of a weighted energy vector [FMS11], as presented in the localization

study in Section 3.1.2. As weighting, $\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}}$ employs the broadband directivity of human hearing again, cf. Section 3.1.2 and Figure 17(b). The regression results for the weighted energy vector $\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}}$ yield R^2 values of 0.75 for both scales.

4.2.2 Lateral Phantom Source Width for Amplitude Panning (Exp. 3)

Conditions

This experiment uses the same conditions as experiment 2. The only difference is the orientation of the subject's head. Instead of looking into the 0° direction, the subjects were facing towards loudspeaker 9 at 33°, cf. gray head symbol in Figure 38. Thus, this experiment evaluates the phantom source width at −33° in the subjective coordinate system.

Results

Figure 40 presents the resulting pooled scale for experiment 3, as well as the median and the corresponding confidence intervals of the 16 individual scales. Again, the correlation between the pooled scale and the median of the individual scales is 0.98.

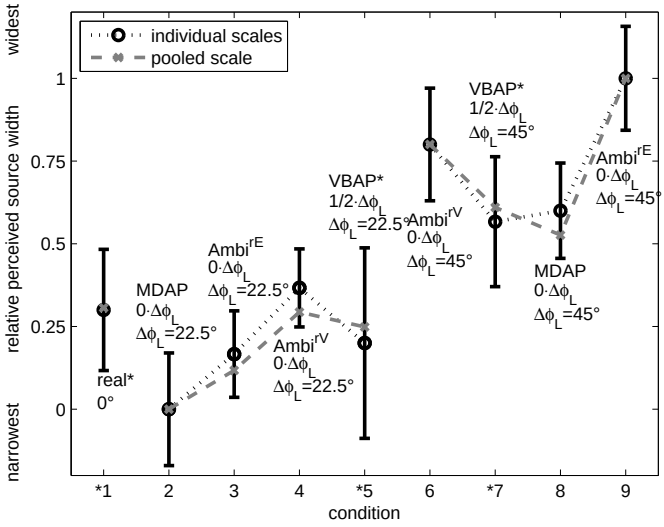


Figure 40: Results of experiment 3 (lateral phantom source width for amplitude panning): Median values and corresponding 95% confidence intervals of individual scales and pooled scale. Median scale and pooled scale are normalized between 0 and 1.

As in the 0° case, the loudspeaker aperture angle $\Delta\phi_L$ has a significant effect on the perceived source width ($p < 0.001$). From all conditions with $\Delta\phi_L = 22.5°$, condition 2 is the only one that is significantly different to condition 1 ($p = 0.008$). Within the conditions for this aperture angle, condition 2 is perceived significantly narrower than condition 4 ($p = 0.003$) and weakly significantly narrower than condition 5 ($p = 0.057$). Within the conditions with $\Delta\phi_L = 45°$, condition 9 is significantly wider than conditions 7 ($p = 0.005$) and 8 ($p < 0.001$). There is a

weakly significant difference between conditions 6 and 8 ($p = 0.055$).
For VBAP, the dependency of the source width on the panning angle is solely significant for the larger loudspeaker aperture angle $\Delta\phi_L = 45°$ (conditions 1/7). This holds true for Ambi^{r_E} (conditions 7/9). For MDAP, there is a weak significance for the dependency on the panning angle in the case of the smaller aperture angle $\Delta\phi_L = 22.5°$ (conditions 2/5). The difference between the two different Ambisonics weightings for panning on the loudspeaker position is not significant for both aperture angles (conditions 3/4 and 6/9).

$\mathbf{R^2}$	$\mathrm{IACC_{E3}}$	$\mathrm{LF^2}$	$\boldsymbol{r}_\mathrm{E}$	$\boldsymbol{r}_\mathrm{E}^\mathrm{w}$
pooled	0.76	0.49	0.54	0.72
median	0.77	0.56	0.54	0.71

Table 12: $\mathrm{R^2}$ values for linear regression of technical measures against the results of experiment 3 (lateral phantom source width for amplitude panning).

For the measurements of $\mathrm{IACC_{E3}}$ and $\mathrm{LF^2}$, the dummy head and the microphone array were also oriented towards loudspeaker 9 at 33°. The same orientation was applied in the directivity weighting of $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$. The $\mathrm{R^2}$ values of the regressions, cf. Table 12, show the same tendencies as in the case of the frontal phantom sources.

4.2.3 Width of VBAP-Panned Phantom Sources (Exp. 4)

Conditions

Experiment 4 evaluates the source width of frontal phantom sources using VBAP at three panning angles with two loudspeakers apertures of $\Delta\phi_L = 22.5°$ and 45°, cf. Table 13. This experiment expands the VBAP-related conditions of experiment 2 by panning to an angle that is one quarter of the loudspeaker aperture angle $\Delta\phi_L$ apart from the loudspeaker position. It was impossible to create a phantom source at a direction of 0° for all conditions on the available loudspeaker setup depicted in Figure 38. The maximum angle between the different phantom source locations was 11°, i.e. for condition 4a. Because a pilot study revealed that subjects had great difficulty in comparing phantom source width when the localization is different, the subjects were told to face each phantom source before evaluating its width. It thus became necessary to allow switching between the conditions in a pair at will.

Results

Figure 41 presents the results of experiment 4. The correlation between the pooled scale and the median of the individual scales is 0.98. As in the previous experiments, the loudspeaker aperture angle $\Delta\phi_L$ has a significant effect on the perceived

condition		* 1a	2a	3a	4a	5a
aperture $\Delta\phi_L$ in °		-	22.5	22.5	45.0	45.0
angle rel. $\Delta\phi_L$		0	1/4	1/2	1/4	1/2
panning method		**each condition uses VBAP**				
number	**ϕ_L in °**	**loudspeaker gains in exp. 4**				
1	-45.0					
2	-22.5				0.94	0.71
3	-11.3		0.95	0.71		
4	-5.6					
5	0.0	1.00				
6	11.3		0.32	0.71		
7	16.8					
8	22.5				0.33	0.71

Table 13: Loudspeaker directions and gains for each condition of experiment 4. Condition 1a also represents VBAP panned on the loudspeaker. All gains $|g_l| < 0.005$ are not shown.

source width ($p = 0.003$). Condition 1a is significantly narrower than every other condition ($p < 0.001$). On the other hand, condition 5a is significantly wider than all other conditions ($p < 0.02$), except for condition 4a. Conditions 2a/3a, 3a/4a, and 4a/5a are not significantly different, whereas the difference between 2a and 4a is weakly significant ($p = 0.052$).

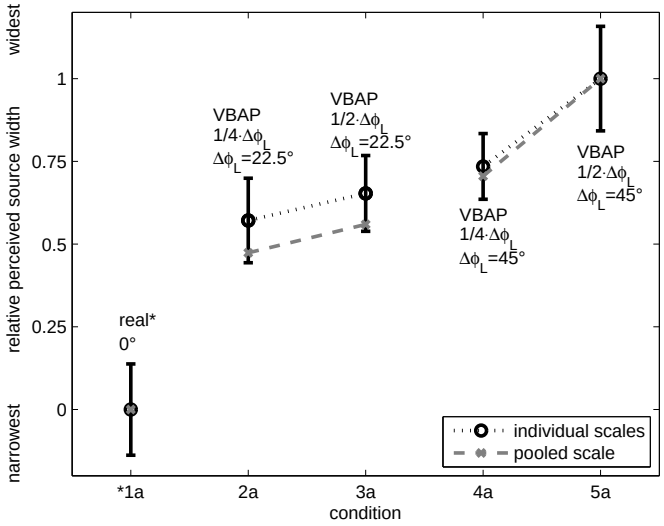


Figure 41: Results of experiment 4 (Width of VBAP-panned phantom sources): Median values and corresponding 95% confidence intervals of individual scales and pooled scale. Median scale and pooled scale are normalized between 0 and 1.

As shown in experiment 2, the source width of VBAP is depending on the panning angle. Panning angles that lie on a loudspeaker yield a narrow source width (condition 1a). The closer the panning angle gets to the middle between the loudspeakers, i.e. $\Delta\phi_L/2$, the wider is the perceived source width. The source width does not differ significantly for panning angles $\Delta\phi_L/2$ and $\Delta\phi_L/4$. These

results are independent of the loudspeaker aperture angle.
By steering into the direction of the phantom source (with a resolution of 5°), the technical measures and predictors also took the different phantom source locations into account. In contrast to the results of experiments 2 and 3, the R^2 values of the regressions in Table 14 show that the $IACC_{E3}$ is the weakest predictor. The R^2 values of $\boldsymbol{r}_E$ and $\boldsymbol{r}_E^w$ are similar, as the loudspeaker pairs are nearly symmetrical. The results for LF^2 are comparable.

$\mathbf{R^2}$	$IACC_{E3}$	LF^2	$\boldsymbol{r}_E$	$\boldsymbol{r}_E^w$
pooled	0.28	0.79	0.78	0.78
median	0.23	0.71	0.67	0.67

Table 14: R^2 values for linear regression of technical measures against the results of experiment 4 (Width of VBAP-panned phantom sources).

4.2.4 Width of Ambisonics-Panned Phantom Sources (Exp. 5)

condition		1b	2b	3b	4b	5b	6b
aperture $\Delta\phi_L$ in °		22.5	22.5	22.5	45.0	45.0	45.0
angle rel. $\Delta\phi_L$		0	1/4	1/2	0	1/4	1/2
panning method		**each condition uses $\mathbf{Ambi^{r_E}}$**					
number	**ϕ_L in °**	**loudspeaker gains in experiment 5**					
1	-45.0	-0.06	-0.04		0.31	0.13	
2	-22.5	0.30	0.12				0.71
3	-11.3			0.71			
4	-5.6						
5	0.0	0.90	0.85		0.89	0.84	
6	11.3			0.71			
7	16.8						
8	22.5	0.30	0.51				0.71
9	33.0						
10	45.0	-0.06	-0.06		0.31	0.52	
11	67.5	0.03	0.02				
12	90.0	-0.02	-0.01		-0.07	-0.07	
13	112.5	0.01	0.01				
14	135.0	-0.01	-0.01		0.04	0.03	
15	157.5	0.01	0.01				
16	180.0	-0.01	-0.01		-0.04	-0.03	
17	-157.5	0.01	0.01				
18	-135.0	-0.01	-0.01		0.04	0.03	
19	-112.5	0.01	0.01				
20	-90.0	-0.02	-0.01		-0.07	-0.04	
21	-67.5	0.03	0.02				

Table 15: Loudspeaker directions and gains for each condition of experiment 5. All gains $|g_l| < 0.005$ are not shown.

Conditions

Experiment 5 repeats experiment 4, but it uses Ambi$^{\text{rE}}$ instead of VBAP, cf. Table 15. Again, the directions of the phantom sources differ up to 11° and the subjects were told to face the phantom sources prior to comparing their widths.

Results

Figure 42 presents the experimental results. The pooled scale and the median of the individual scales correlate with 0.99. The loudspeaker aperture angle $\Delta\phi_L$ still has a significant effect on the perceived source width ($p < 0.001$). Conditions 1b, 2b, and 3b are not significantly different. The difference between conditions 4b and 5b is weakly significant ($p = 0.06$), and the difference between conditions 4b and 6b is significant ($p = 0.002$). Conditions 5b and 6b are not significantly different.

The results for panning between and on a loudspeaker agree with the findings from experiment 2. For the smaller loudspeaker aperture $\Delta\phi_L = 22.5°$, the source width of Ambi$^{\text{rE}}$ is independent of the panning angle. For an aperture angle of $\Delta\phi_L = 45°$, the source width for panning angles on a loudspeaker is significantly wider compared to panning angles exactly in middle between the loudspeakers. Despite this difference, panning in the middle and quarter between the loudspeakers yields no significant change of the source width.

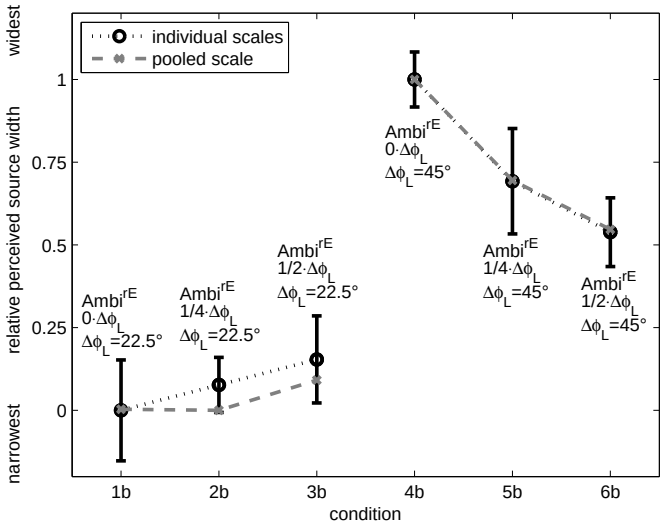


Figure 42: Results of experiment 5 (Width of Ambisonics-panned phantom sources): Median values and corresponding 95% confidence intervals of individual scales and pooled scale. Median scale and pooled scale are normalized between 0 and 1.

The weighted energy vector is the best predictor for experiment 5, cf. Table 16. Omission of the weighting reduces the determination of the energy vector. LF^2 and IACC_{E3} yield comparable results.

$\mathbf{R^2}$	$\mathrm{IACC_{E3}}$	$\mathrm{LF^2}$	$\boldsymbol{r}_\mathrm{E}$	$\boldsymbol{r}_\mathrm{E}^\mathrm{w}$
pooled	0.69	0.80	0.87	0.92
median	0.68	0.77	0.85	0.89

Table 16: $\mathrm{R^2}$ values for linear regression of technical measures against the results of experiment 5 (Width of Ambisonics-panned phantom sources).

4.2.5 Discussion

In experiment 2, the perceived source width of VBAP significantly depends on the panning angle: the width is smaller for panning angles on the loudspeakers. The same tendency can be found for MDAP. However, the dependency is weaker. These results agree with the findings of Pulkki [Pul99]. On the other hand, $\mathrm{Ambi^{r_E}}$ yields the widest sources for panning angles on the loudspeakers. Nevertheless, a clearly noticeable panning dependence is only found for the larger loudspeaker aperture angle of $\Delta\phi_L = 45°$. Only for this aperture angle, the difference between the two Ambisonics weightings is significant.

Experiments 4 and 5 showed for VBAP and $\mathrm{Ambi^{r_E}}$ and both loudspeaker aperture angles that the trend of the source width can be interpolated between panning on a loudspeaker and half way between a loudspeaker pair.

Independent of the panning method, a larger loudspeaker aperture angle yields a wider phantom source in all experiments. As the investigated conditions with the larger aperture angle also involve a lower Ambisonic order, lower-order phantom sources are perceived wider in this study. This dependency agrees with the findings for first and second order Ambisonics in [MWCQ99b].

The different head orientation in experiment 3 yields larger intersubjective variation in the results and therefore less significant differences. This agrees with the large intersubjective differences in the localization of lateral phantom sources.

The $\mathrm{IACC_{E3}}$ is a good predictor for the results from most of the experiments. However, it is the weakest predictor for experiment 4. This agrees with the results from Section 4.1, where the source width of the two-loudspeaker conditions was poorly predicted, while the prediction of the conditions with three loudspeakers was better.

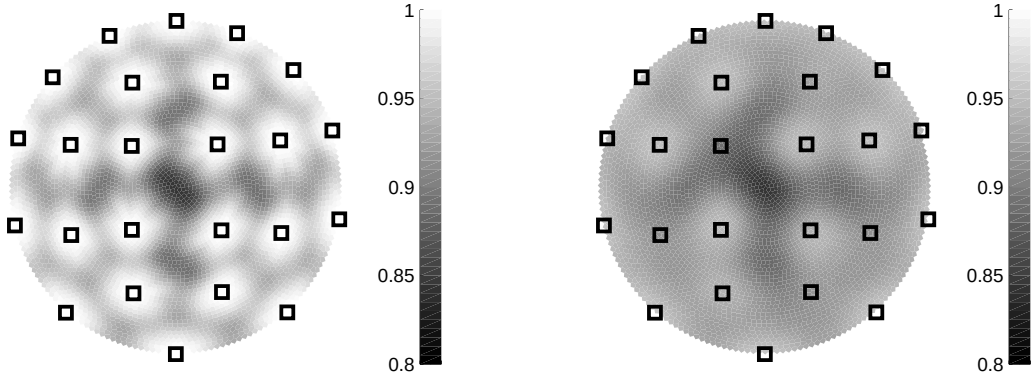
The average performance of $\mathrm{LF^2}$ is slightly better. The incorporation of direction-dependent weighting in $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$ increases the $\mathrm{R^2}$ by 0.2 in comparison to the unweighted $\boldsymbol{r}_\mathrm{E}$. Moreover, the $\mathrm{R^2}$ values of $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$ never fall below 0.72 (for the pooled scales). Thus, the $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$ predictor seems to offer the most robust prediction.

4.3 Further Studies on Phantom Source Width

4.3.1 Phantom Source Width in 3D

The length of the (yet unweighted) energy vector $\boldsymbol{r}_\mathrm{E}$ is also used as a simple estimator for the energy spread for panning methods in 3D [ZFS10, ZPN12, ZF12]. The estimated spread corresponds with the source width in the listening experience. Still, formal listening tests about the suitability of the measure for phantom source width in 3D have to be performed.

As an example, Figure 43(a) shows the predicted energy spread using VBAP for different panning directions on the hemispherical loudspeaker arrangement described in Table 8. The view is a projection of the upper hemisphere onto a circle, so that the center of the circle represents the north pole of the hemispherical arrangement. The energy is focused whenever panning directions are aligned with loudspeakers (black squares), and it spreads for panning directions between the loudspeakers just as in the 2D case, cf. Section 4.2. Figure 43(b) shows the increased smoothness of Ambi$^{\mathrm{r_E}}$ panning. This example uses the so-called AllRAD decoder design [ZF12] that is optimized for irregular arrangements covering only parts of a whole sphere.



(a) VBAP (b) AllRAD 5$^{\mathrm{th}}$ order Ambi$^{\mathrm{r_E}}$

Figure 43: $|\boldsymbol{r}_\mathrm{E}|$ for panning on the upper hemisphere, loudspeaker positions are marked by black squares. Figure adapted from [ZF12].

4.3.2 Phantom Source Widening using Stereo-Decorrelation

One way of increasing the spatial extent of a phantom source is the decorrelation of the loudspeaker signals, typically of a loudspeaker pair [Sch58, Orb70, Ger93, Ken95, BK04, Pot06]. It is desirable that a suitable pair of decorrelation filters does not introduce additional changes in the signals, such as coloration.
An efficient filter structure presented in [ZF13] achieves this by frequency-dependent phase or amplitude panning. It controls the widening by influencing the inter-channel cross correlation coefficient (ICCC). The ICCC can be also calculated from the filter impulse responses like the IACC using Eqs. (15) and (16). A listening test [ZFMS11] using the phase-based method on a standard $\pm 30°$ loudspeaker pair yielded a high correlation between the phantom source width and the $\mathrm{IACC_{E3}}$ measures ($\mathrm{R}^2 = 0.96$). The decorrelation of the loudspeaker signals causes a decorrelation of the ear signals linearly related to ICCC within the tested range from 0.6 to 1.
In [FZ13], we studied the applicability of the simpler LF^2 and $\boldsymbol{r}_\mathrm{E}$ predictors for the prediction of phantom source widening by decorrelation. Both measures are independent of the ICCC, as they are only consider the directional distribution of the wide-band signal energy and do not take the phase distribution into account. However, both measures can be extended by a factor that accounts for the correlation of the loudspeaker signals. A lateral energy fraction measure $\mathrm{LF}^2_\mathrm{corr}$ can be defined as

$$\mathrm{LF}^2_\mathrm{corr} = \mathrm{LF}^2 - \mathrm{ILCC} + 1, \tag{23}$$

where the ILCC is the inter-loudspeaker cross correlation coefficient, i.e., the measured coherence between the impulse responses $h_\mathrm{o,LS1}(t)$ and $h_\mathrm{o,LS2}(t)$ of each filtered loudspeaker signal to an omnidirectional microphone. The ILCC is calculated like the IACC and the ICCC.
Similarly, the $\boldsymbol{r}_\mathrm{E}$ estimator that should not require measurements, can be extended by the ICCC

$$\boldsymbol{r}_\mathrm{E,corr} = \boldsymbol{r}_\mathrm{E} \cdot \mathrm{ICCC}. \tag{24}$$

The extended measures $1 - \mathrm{LF}^2_\mathrm{corr}$ and $||\boldsymbol{r}_\mathrm{E,corr}||$ are linearly related to $\mathrm{IACC_{E3}}$ and ICCC. Still, both predictors do not require dummy head measurements as the $\mathrm{IACC_{E3}}$ would.

4.4 Conclusion on Phantom Source Width

This chapter studied the perception of phantom source width created by amplitude panning in the horizontal plane and its prediction. The first experiment showed that a phantom source is perceived wider than a single loudspeaker. When a phantom source is created between two symmetrically arranged loudspeakers, phantom source width increases with an increasing aperture angle of the loudspeaker pair. This holds true for arrangements with an additional center loudspeaker between the loudspeaker pair. In comparison to the loudspeaker pair, the additional center loudspeaker decreases the phantom source width. This can be explained by a decrease in the relative share of lateral sound.

The experimental results were compared to $\mathrm{IACC_{E3}}$ and LF measures, which are established predictors for source width in room acoustics using dummy head and microphone array recordings, respectively. The $\mathrm{IACC_{E3}}$ turned out to be a fair predictor of phantom source width. In contrast, the $\mathrm{LF^{ISO}}$ according to the ISO norm is not suitable at all in this case. The poor predictability of the experimental results by the $\mathrm{LF^{ISO}}$ is caused by simultaneous sound incidence from multiple, symmetrically arranged directions that yield signal cancellation at the figure-of-eight microphone. Therefore, a modified $\mathrm{LF^2}$ measure has been introduced that uses a energetic superposition of the loudspeaker signals by sequential measurements. This new measure is highly correlated to the experimental results and can be well approximated by simultaneous measurement at multiple positions. An even better correlation is achieved by the length of the energy vector $\boldsymbol{r}_\mathrm{E}$ that is computed solely from the loudspeaker positions and gains and thus does not require measurements at all.

Further experiments extended the study to frontal amplitude-panning scenarios using VABP, MDAP, $\mathrm{Ambi^{r_E}}$, and $\mathrm{Ambi^{r_V}}$ on a regular ring of 8 and 16 loudspeakers. As in the previous experiment, a larger loudspeaker aperture angle, i.e. a smaller number of loudspeakers, yields a larger phantom source width. In comparison to the other panning methods, VBAP results in larger source width variation when panning on or half way between two loudspeakers. For lateral directions, the assessment of phantom source width is more difficult and the variation in the experimental results increases.

The results of the experiments can also be predicted by $\mathrm{IACC_{E3}}$ and $\mathrm{LF^2}$, but the length of the basic $\boldsymbol{r}_\mathrm{E}$ is not suitable this time. In comparison to the experimental results, the prediction overestimates the influence of the loudspeakers behind the listener, and it underestimates the influence of the lateral loudspeakers. The proposed weighted energy vector $\boldsymbol{r}_\mathrm{E}^\mathrm{w}$ considers a direction-dependent amplitude-weighting of the loudspeaker gains. A suitable weighting can be found in the

directivity of human hearing. The length of $\boldsymbol{r}_{\mathrm{E}}^{\mathrm{w}}$ yields a robust prediction of the experimental results. The vector model is also proposed as a simple predictor for phantom source width in 3D panning and it can incorporate decorrelation of loudspeaker signals for the prediction of phantom source widening.

Chapter V

PHANTOM SOURCE COLORATION

Amplitude-panning methods employ amplitude differences between the loudspeakers to control the direction of a phantom source. Besides the direction, other attributes of the phantom source are modified, unintentionally. The modification of phantom source width was presented in the previous chapter. This chapter discusses coloration, i.e. changes in the timbre of the phantom source. In [RZKB05], we find the statement about the overall quality of a spatial sound rendering system being composed of timbral fidelity by 70% and of spatial fidelity by 30%. This finding emphasizes the importance of timbral fidelity, i.e., it is paramount for a sound rendering system to avoid any coloration.

In most studies, coloration has been assessed by scaling of verbal descriptors and attribute ratings [Rum02, ZK01, CW07, GK04]. Methods such as multidimensional scaling [MRZ08] or the repertory grid technique [BR99] are helpful to determine the dimensionality and components of coloration. An assessment of coloration is difficult due to the verbal descriptors that are often difficult to define consistently. In general, the coloration of a sound rendering system can often be equalized statically as long as the system does not produce annoying, time-variant coloration. Thus, this chapter focuses on coloration changes during panning. This is done by a listening experiment. As predictor, a binaural measure is applied that is based on third-octave band levels.

Section 5.1 presents the simplification of the binaural coloration model by Ono [OPK01, OPK02]. First, the model is applied to a single real source and can be used to explain binaural decoloration [Sal95, Brü01] by simple signal processing means. The application is then extended to frontal phantom sources created by two and three simultaneously playing loudspeakers with the same gain. The resulting notches in the third-octave band spectrum can be estimated by comb filters that are caused by the inter-aural delay time between the loudspeakers. Section 5.2 investigates the dependency of coloration on the panning angle using VBAP, MDAP, and Ambisonics. A listening experiment in Section 5.3 studies the coloration of moving phantom sources in the central and an off-center listening position. The experimental results can be explained by the incremental changes in the third-octave band spectrum during panning. Finally, a link between coloration and the variation in the length of the energy vector, as well as between coloration and the number of active loudspeakers, is established in Section 5.4.

5.1 Coloration using 1, 2, 3 Loudspeakers

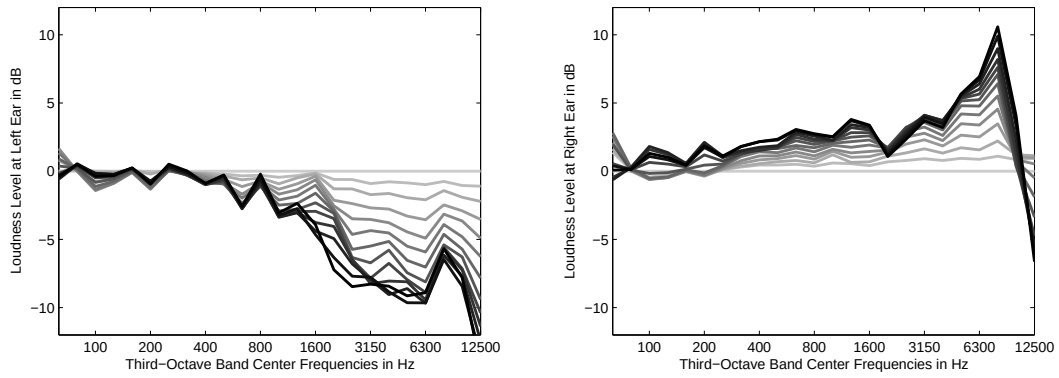
5.1.1 Composite Loudness Level

The Composite Loudness Level (CLL) is proposed as a model for the perceived coloration in [OPK01, OPK02] and is basically a binaural loudness level spectrum. The binaural input signal is fed into a model of middle ear, cochlea, and auditory nerve [PKH99, PK01]. The cochlea is modeled by a filter bank with 42 bands with center frequencies and bandwidth according to the ERB scale [MPG90]. For each ear and frequency band, the loudness is calculated according to [ZF90] and summed up. Results in [OPK01, OPK02] show that the perceived timbre corresponds to the sum of both loudness spectra.
The present work simplifies the CLL by excluding the models for middle ear, cochlea and auditory nerve. Furthermore, it uses third-octave bands according to ANSI S1.6-1984 [ANS06]. In each frequency band b, the loudness N_b (in sone) is calculated from the intensity levels I_b (in dB) by $\mathrm{N}_b = 2^{\frac{\mathrm{I}_b - 40}{10}}$. Note that this formula is originally used to translate loudness level in phon to loudness in sone [ZF90]. In each frequency band, the loudness for both ears $\mathrm{N}_{b,left}$ and $\mathrm{N}_{b,right}$ is summed up and translated to loudness levels CLL_b in phon/dB by $\mathrm{CLL}_b = 40 + 10\log_2(\mathrm{N}_{b,left} + \mathrm{N}_{b,right})$.
Due to the limited frequency range of the loudspeakers, the first 4 bands are omitted resulting in 24 bands from 63Hz to 12.5kHz. The third-octave band levels are calculated from the first 80ms of the binaural impulse responses. The same time window was used for the binaural localization model in Section 3.2.1 and the width measures in Section 4.1.
Studies about loudspeaker and headphone equalization show that differences in third-octave band levels of less than 1dB are inaudible by most listeners [RS87, KPJH99]. This criterion is also used in this work for the perception of coloration, i.e., differences between CLL spectra of less than 1dB are assumed to be inaudible.

5.1.2 Coloration using 1 Loudspeaker

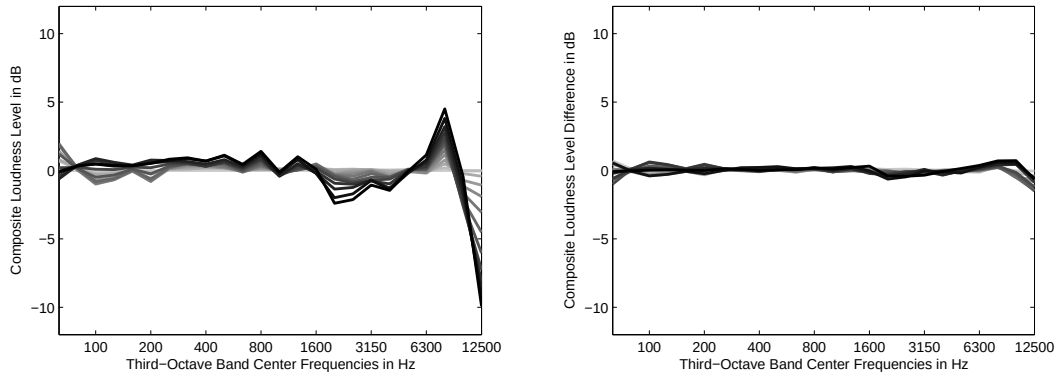
Figures 44(a) and 44(b) show the third-octave band spectra at the left and right ear for a single loudspeaker that is rotated around the dummy head in the listening room from 0° to −60° (clockwise) in 5° steps. Obviously, the left ear signal has a higher level caused by the ILD. The variation of the spectra exceed the 1dB limit and would yield a coloration if the ears were treated separately.
However, experiments show that in case of a slow movement of the loudspeaker the listeners can hardly perceive coloration [The80]. The phenomenon that the coloration is suppressed when listening with both ears is called binaural decoloration [Sal95, Brü01]. Figure 45(a) indicates that the CLL spectrum is flatter

compared to the spectra of each ear, cf. Figures 44(a) and 44(b).
However, the spectrum for the $-60°$ direction largely deviates from the spectrum for the $0°$ direction at higher frequencies. The above mentioned experiments [The80] did not compare the coloration of different discrete directions directly but used a slow movement. Thus, it seems that rather the incremental spectral differences are important than the absolute changes. Figure 45(b) shows the incremental spectral differences in steps of $5°$. The differences are within the 1dB limit for all increments. When increasing the step size to $10°$ and $15°$, the upper frequency limit that yields differences below 1dB decreases to 10kHz and 8kHz, respectively.
The results of the incremental spectral differences agree with the binaural decoloration of slowly moving real sound sources [The80]. Thus, the incremental differences using a step size of $5°$ are employed as a measure for coloration when comparing different panning methods in Section 5.3.



(a) spectra at left ear (b) spectra at right ear

Figure 44: Loudness Level spectra at left and right ear for a single loudspeaker at positions between $0°$ (light gray) and $-60°$ (black) in $5°$ steps, normalized to $0°$.



(a) spectra, normalized to $0°$ (b) incremental spectral differences

Figure 45: Composite Loudness Level spectra and incremental differences for a single loudspeaker at positions between $0°$ (light gray) and $-60°$ (black) in $5°$ steps.

5.1.3 Coloration using 2 Loudspeakers

This paragraph studies phantom source coloration when using 2 simultaneously active loudspeakers with different aperture angles $\Delta\phi_L$ between them. All loudspeaker pairs are symmetrically arranged around the 0° direction and each loudspeaker in a pair is fed by the same signal with the same gain. The coloration is estimated by the CLL spectra. These are computed from the same impulse responses that were measured for the prediction of source width in Section 4.1.

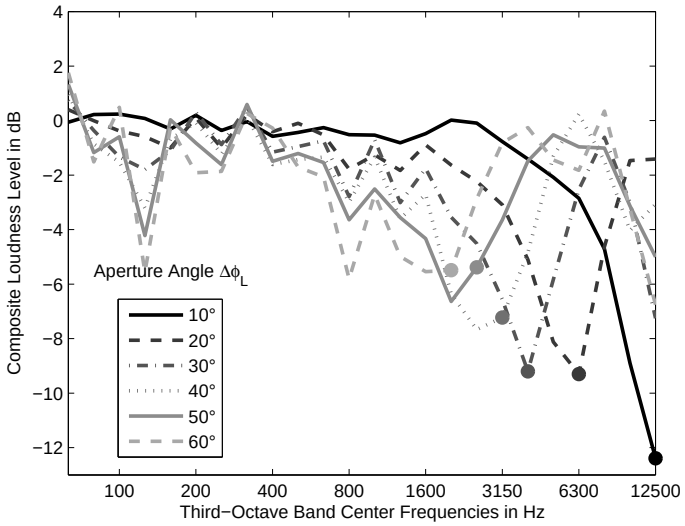


Figure 46: 0°-normalized Composite Loudness Level spectra for symmetrically arranged loudspeaker pairs with different aperture angles $\Delta\phi_L$. Points indicate the first notch caused by a comb filter with the corresponding ITD values (frequencies quantized by third-octave band center frequencies).

Figure 46 shows the CLL spectra for aperture angles from 10° to 60°. The spectra are normalized to the CLL spectrum of a single loudspeaker at 0°. The notch at 125Hz is most probably caused by floor reflections for lateral loudspeaker positions. More importantly, there is a prominent high-frequency notch whose frequency and depth depends on the loudspeaker spacing. Both parameters decrease for larger aperture angles. In order to explain this dependency, Figure 47 presents a sketch of the binaural impulse responses of each loudspeaker and their superposition for a loudspeaker pair with an aperture angle of 10° (loudspeakers at ±5°). For increasing aperture angles, both the ILD and the ITD increase. Thus, the time-delay of the second peak in the superimposed impulse response increases, and its level decreases for enlarged aperture angles of the loudspeaker pair. The resulting comb filter decreases its frequency and depth. The points in Figure 46 show that this simplified ITD-based comb filter model can estimate the first notch of the computed CLL spectra.

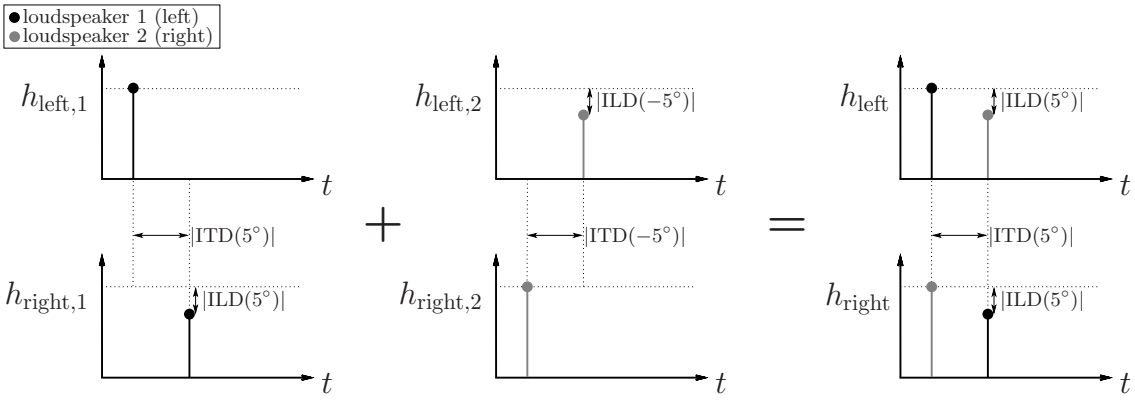


Figure 47: Sketch of impulse responses of each loudspeaker 1, 2 at each ear ($h_{\text{left},1}$, $h_{\text{right},1}$), ($h_{\text{left},2}$, $h_{\text{right},2}$), and sumperimposed impulse responses (h_{left}, h_{right}). Loudspeakers are located at ±5°.

5.1.4 Coloration using 3 Loudspeakers

The above-mentioned study is extended to three simultaneously active loudspeakers by employing the same setup, except for the additional center loudspeaker at 0° that has the same gain as each of the loudspeakers in the pair. Again, the binaural impulse responses from Section 4.1 are used.

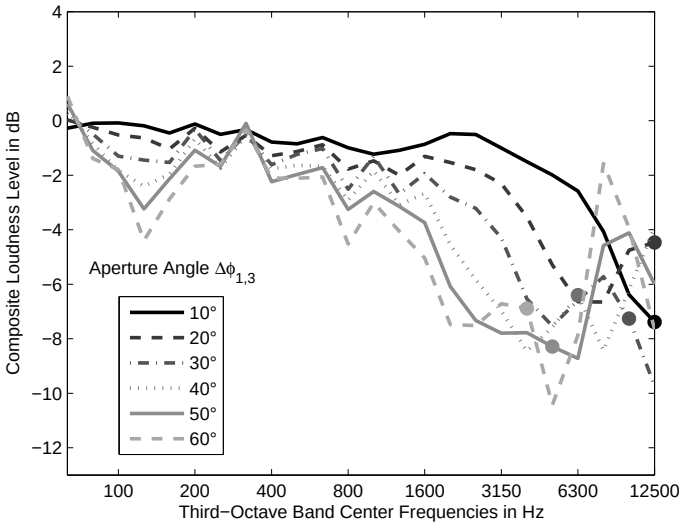


Figure 48: 0°-normalized Composite Loudness Level spectra for symmetrically arranged loudspeaker triplets with different aperture angles $\Delta\phi_{1,3}$ between the two outmost loudspeakers (1,3). Points indicate the first notch caused by a comb filter with the corresponding ITD values (frequencies quantized by third-octave band center frequencies).

The resulting CLL spectra from 3 loudspeakers are shown in Figure 48 for different aperture angles $\Delta\phi_{1,3}$ between the two outmost loudspeakers 1 and 3, i.e. the loudspeaker pair. Comparing the results to the 2-loudspeaker case reveals that

the addition of the center loudspeaker increases the frequency of the prominent notch. Note that the notch frequencies for aperture angles $\leq 20°$ are ≥ 12.5kHz.

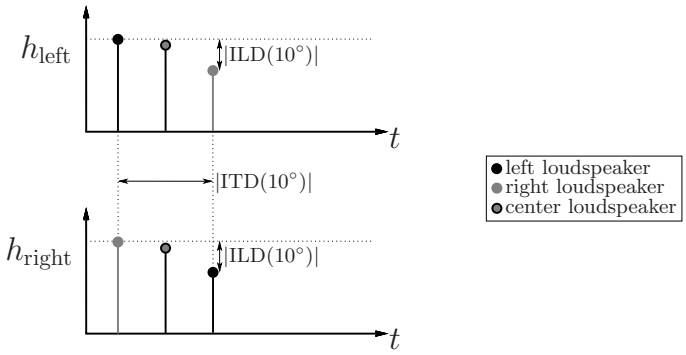


Figure 49: Sketch of the superimposed impulse responses h_{left}, h_{right} at each ear for a loudspeaker pair with an aperture angle of 20° (loudspeakers at ±10°) and a simultaneously active center loudspeaker (at 0°).

The higher frequencies of the notches when using 3 loudspeakers can be explained by the impulse responses, cf. Figure 49. Just as in the 2-loudspeaker case, there are two peaks that are caused by the ipsilateral and contralateral loudspeaker of the symmetrical loudspeaker pair. The level difference and time delay between both peaks depend on the aperture angle of the loudspeaker pair. However, there is an additional peak without inter-aural time delay and level difference that is caused by the center loudspeaker. Nevertheless, the time delay between the additional peak and the ipsilateral peak is approximately half the inter-aural time-delay of the loudspeaker pair and is thus depending on the aperture angle of the loudspeaker pair. Summing up, this results in 3 approximately equidistant peaks in the superimposed impulse responses h_{left} and h_{left}, where the temporal distance and the level of the peaks is depending of the aperture angle $\Delta\phi_{1,3}$ of the loudspeaker pair. In comparison to the loudspeaker pair without the center loudspeaker, the addition of the center loudspeaker yields a notch at twice the frequency. The hereby estimated notch frequencies are indicated as points in Figure 48 and are similar to the notches of the calculated CLL spectra.

5.2 Coloration using Amplitude Panning with Multiple Loudspeakers

The generic cases of the previous section are extended by studying the coloration of typical amplitude-panning scenarios. The study employs VBAP, MDAP (B = 10, $\phi_{\mathrm{MDAP}} = {}^1\!/_2\Delta\phi_L$), and Ambi$^{\mathrm{rE}}$ on a ring with 16 (7^{th} order Ambisonics) and 8 (3^{rd} order Ambisonics) equidistantly arranged loudspeakers (aperture angles $\Delta\phi_L$ of 22.5° and 45°, respectively) and compares panning angles on and exactly between the loudspeakers, as it was used in Section 4.2 about source width. The loudspeaker arrangement is rotated in a way that provides the same 0° phantom source location in all cases independent of the panning angle.

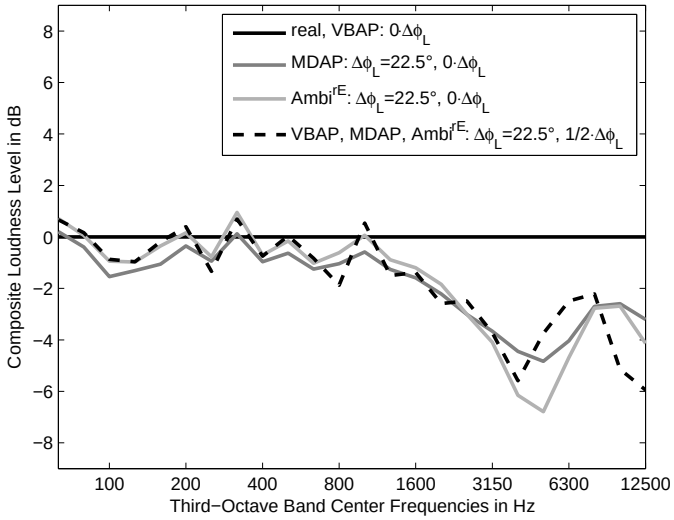


Figure 50: Composite Loudness Level spectra for different panning angles and panning methods using a loudspeaker aperture angle of $\Delta\phi_L = 22.5°$.

Figure 50 presents the resulting CLL spectra for different panning methods and angles using a loudspeaker aperture angle of $\Delta\phi_L = 22.5°$. The spectra are normalized to the spectrum of a single loudspeaker at 0°, i.e. a real sound source. VBAP, MDAP, and Ambi$^{\mathrm{rE}}$ for panning between the loudspeakers, as well as MDAP and Ambi$^{\mathrm{rE}}$ panned on a loudspeaker yield similar CLL spectra with strong deviations from the spectrum of the real loudspeaker. All these cases show a prominent notch at about 5kHz which fits to the estimations from Figure 46 for an aperture angle of 20° to 30°. VBAP panned on the loudspeaker employs only a single loudspeaker and obviously yields a flat spectrum. However, this means that VBAP results in the greatest coloration difference between panning on and between the loudspeakers, whereas the coloration of MDAP and Ambi$^{\mathrm{rE}}$ is less dependent on the panning angle. It is desirable for smooth moving sources with least coloration changes that this dependency is as small as possible.

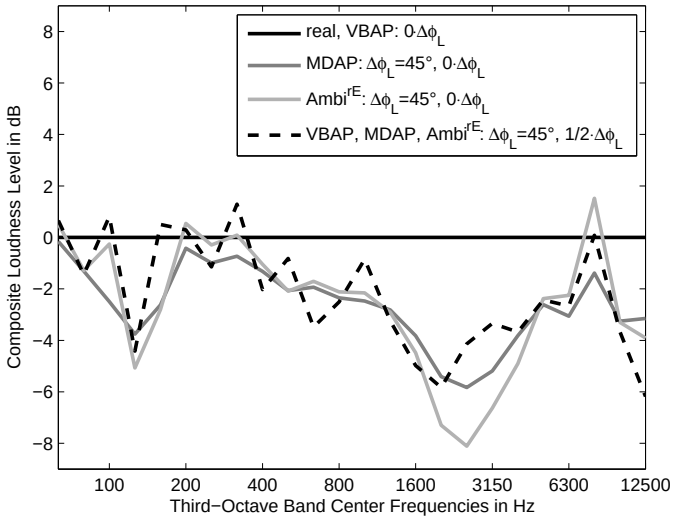


Figure 51: Composite Loudness Level spectra for different panning angles and panning methods using a loudspeaker aperture angle of $\Delta\phi_L = 45°$.

Figure 51 presents the resulting CLL spectra for the same cases but using an enlarged loudspeaker aperture angle of 45°. Due to the larger aperture angle, the prominent notch is shifted to about 2kHz, similarly to the notches for 40° and 50° in Figure 46. Still, the same tendencies as in the 22.5° case hold: The coloration of MDAP and Ambi$^{\mathrm{rE}}$ is less dependent on the panning angle than VBAP.

To summarize, there is a prominent notch for VBAP, MDAP, and Ambi$^{\mathrm{rE}}$ for panning between the loudspeakers, as well as MDAP and Ambi$^{\mathrm{rE}}$ for panning on the loudspeaker. The frequency of the notch is depending on the loudspeaker aperture angle, according to the previous estimations using 2 or 3 loudspeakers. For a fixed aperture angle, the CLL spectra MDAP and Ambi$^{\mathrm{rE}}$ are similar and independent of the panning angle. This global, panning-independent deviations from a flat spectrum can therefore be equalized statically. This does not apply to VBAP: When panning to the direction of a single loudspeaker, VBAP employs only this single loudspeaker. This results in no comb filtering compared to panning between 2 loudspeakers. VBAP yields the most panning-dependent coloration because it alternately uses 1 or 2 loudspeakers. This coloration change is expected to become audible whenever a phantom source is moving around the listener.

5.3 Coloration of Moving Phantom Sources

This section studies the perceived coloration change of moving phantom sources panned all around the listener in a listening experiment. The experiment employs a regular arrangement with 8 and 16 loudspeakers, with loudspeaker aperture angles $\Delta\phi_L$ of 45° and 22.5°, respectively, cf. Figure 52.

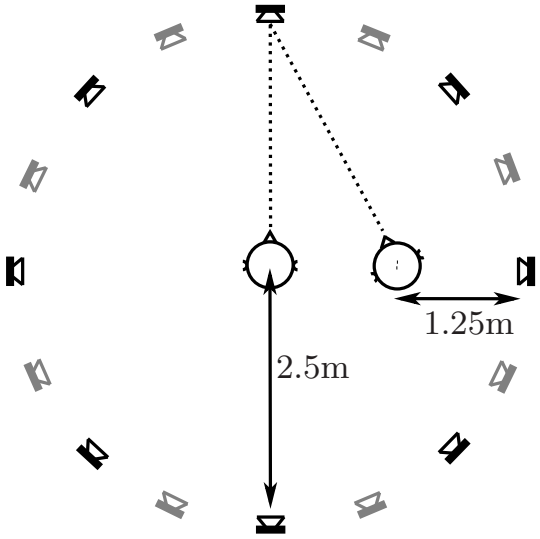


Figure 52: Experimental Setup using 8 (black) and 16 (black+gray) loudspeakers; central and off-center listening positions.

It employs VBAP, MDAP, Ambi$^{\text{rE}}$, and Ambi$^{\text{rV}}$ on both arrangements. The Ambisonics conditions make use of the highest possible orders, i.e. a maximum order of 3 for 8 loudspeakers and 7 for 16 loudspeakers. MDAP is applied with B = 10 panning directions within $\phi_{\text{MDAP}} = {}^1\!/_2\Delta\phi_L$. One of the conditions was repeated in order to verify the intrarater reliability of each subject. This results in a total of 9 conditions that were randomly arranged for each trial. There were 4 trials in the experiment, two at the central listening position and two at the off-center position. In each position, the movement of the phantom source was presented clockwise and counter-clockwise. In either way, the rotation speed was 0.1° per ms using an interpolation time of 1ms. This results in 3.6s for whole 360° movement around the listener. In preliminary test, this speed was found to be a good choice in terms of condition discriminability, subject annoyance, and duration of the experiment. The stimulus was continuous pink noise at 65dB(A). The listeners were allowed to switch between the different conditions at will during the source movement by pressing buttons on a touch screen, cf. Figure 53. The touch screen was also used for the judgment of the coloration changes on a continuous scale from "imperceptible" to "very intense" (the test used the German terms „nicht wahrnehmbar "and „sehr deutlich"). The user interface also provided the possibility of ordering the judgments for better comparison of similar conditions.

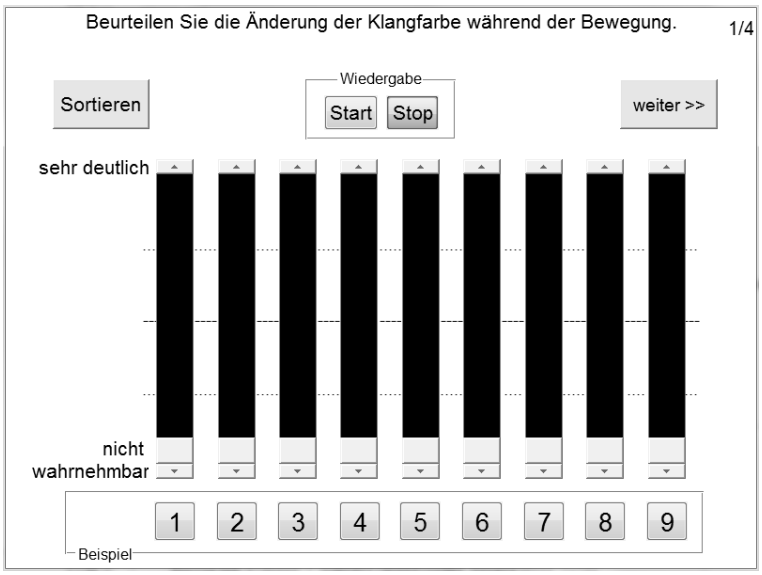


Figure 53: Graphical user interface of the experiment, presented on a touch screen.

A total of 14 subjects participated in the experiment. Due to the high intrarater reliability, no subjects were excluded from the results.

5.3.1 Central Listening Position

The median values for clockwise and counter-clockwise movements were nearly identical and correlated with 0.99 for the central listening position. Therefore, responses for opposite movements were considered as repetitions of the same conditions. This yields 28 answers for each condition.

Figure 54 presents the medians and their confidence intervals for the 8 conditions evaluated at the central listening position. Within the group of the loudspeaker aperture $\Delta\phi_L = 45°$, Ambi$^{\mathrm{rV}}$ yields significantly more coloration than MDAP and Ambi$^{\mathrm{rE}}$ ($p < 0.006$), and its difference to VBAP is not significant ($p = 0.30$). The difference between MDAP and Ambi$^{\mathrm{rE}}$ is insignificant ($p = 0.41$).

Within the group of the $\Delta\phi_L = 22.5°$ aperture angle, all differences are significant ($p < 0.003$). VBAP causes the most intense coloration changes, followed by Ambi$^{\mathrm{rV}}$, and MDAP. The least changes are perceived when using Ambi$^{\mathrm{rE}}$.

The group with the larger aperture produced significantly less coloration than the group with the smaller aperture ($p < 0.001$). In detail, the increase of the coloration with the increasing number of loudspeakers is significant for all panning methods ($p < 0.001$), except for the weakly significance for Ambi$^{\mathrm{rE}}$ ($p < 0.062$). The dependency on the number of loudspeakers might be due to the sensitivity of fluctuation strength. For broadband noise, the maximum of the sensitivity was determined for an amplitude modulation frequency of 4Hz [ZF90]. The sensitivity decreases towards higher and lower frequencies. Considering the rotation speed of

the phantom source in the experiment, the conditions using 8 loudspeakers result in a modulation frequency of 2.2Hz, while the frequency for the 16-loudspeaker conditions yields 4.4Hz.

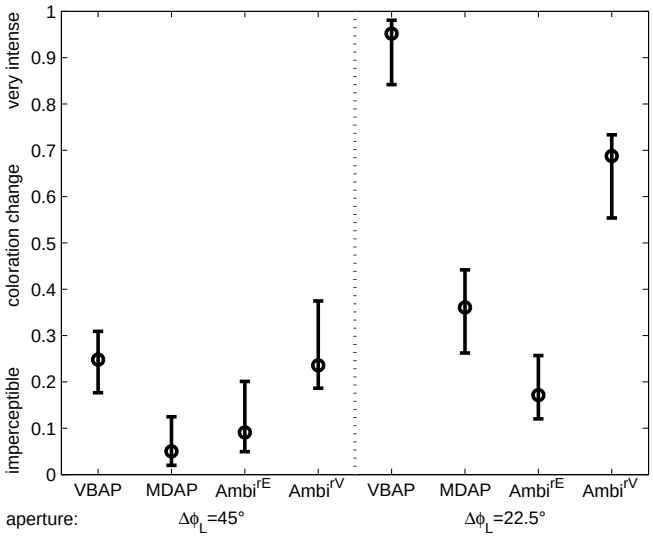
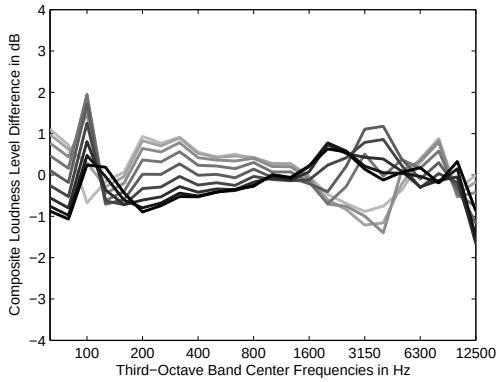


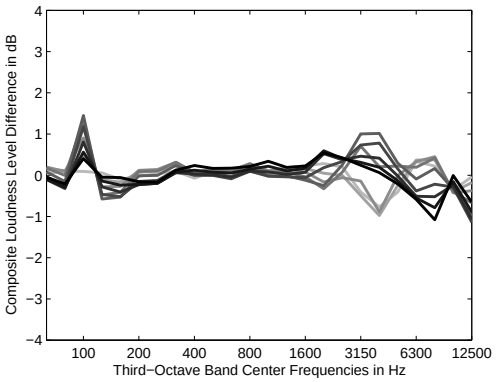
Figure 54: Medians and corresponding 95% confidence intervals of the coloration changes at the central listening position during panning for different amplitude panning methods and loudspeaker apertures.

Figures 55(a) to 55(h) present the incremental changes in the CLL spectra when panning from 0° to 45° in 5° steps. Changes have a larger impact when using 16 loudspeakers. The stronger spectral changes can be explained by the reduced inter-aural time-delay differences for smaller loudspeaker aperture angles, which cause deeper comb filters, cf. Figure 46. The stronger perceived coloration for these conditions can be explained by the perception of fluctuation strength that increases with increasing modulation depth [ZF90].

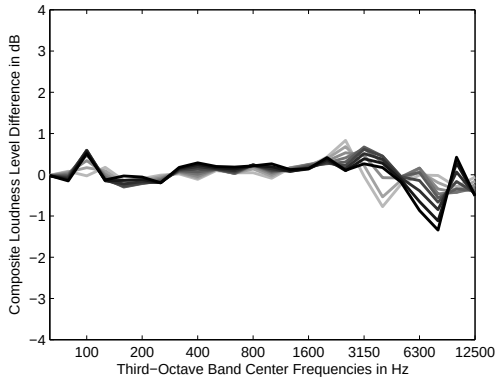
In order to further compare the measurements to the experimental results, it is desirable to transform the measurements into a single value for each of the 8 conditions. This is done by first taking the maximum of each incremental CLL change. As there are 9 incremental steps of 5°, this results in 9 values for each condition. From these 9 values, the maximum value is selected. The thereby obtained values correspond to the strongest spectral changes during panning and correlate with 0.90 to the median results of the listening experiment.



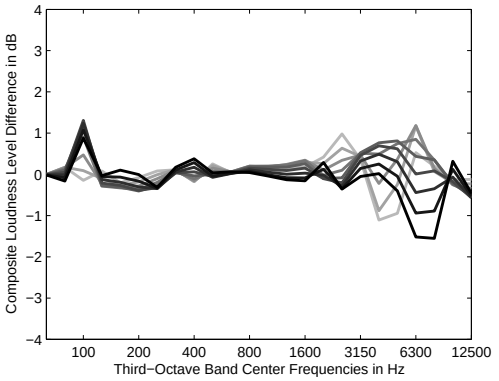
(a) VBAP with 8 loudspeakers



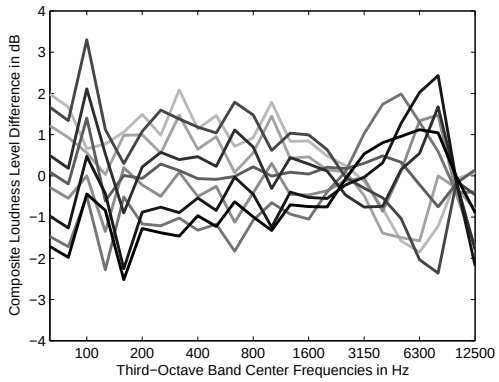
(b) MDAP with 8 loudspeakers



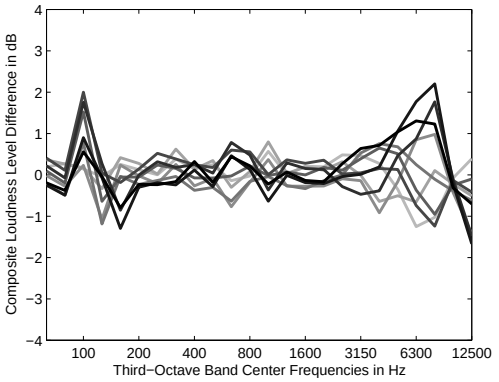
(c) Ambi^{r_E} with 8 loudspeakers



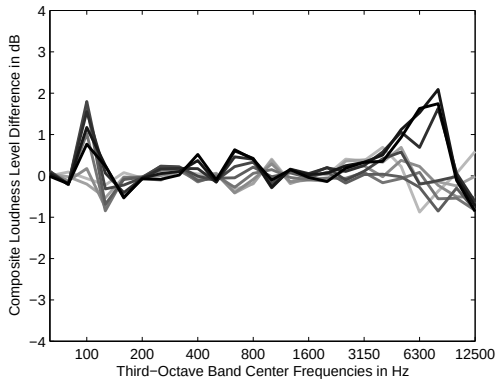
(d) Ambi^{r_V} with 8 loudspeakers



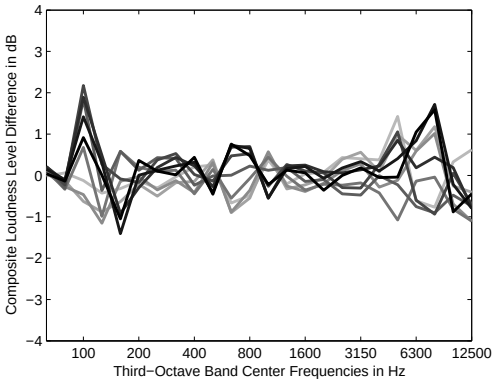
(e) VBAP with 16 loudspeakers



(f) MDAP with 16 loudspeakers



(g) Ambi^{r_E} with 16 loudspeakers



(h) Ambi^{r_V} with 16 loudspeakers

Figure 55: Incremental differences in the Composite Loudness Level Spectra for panning from 0° to 45° in 5° steps at the central listening position.

5.3.2 Off-Center Listening Position

The same experiment was done at the off-center listening position, cf. Figure 52. Again, the correlation between the clockwise and counter-clockwise movement correlate with a high level of 0.98 and are therefore not treated independently.

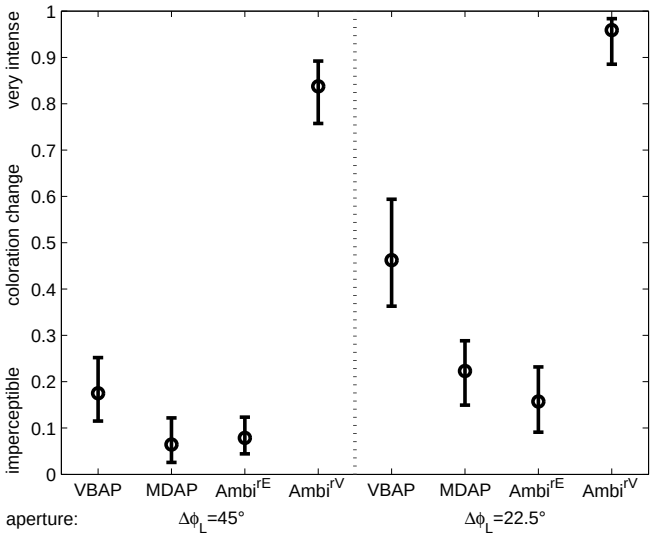
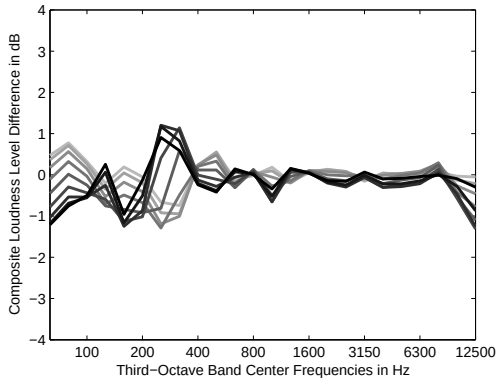
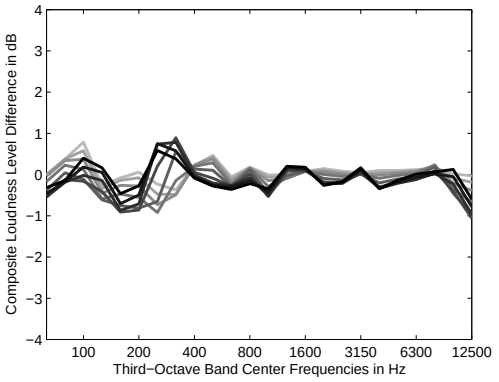


Figure 56: Medians and corresponding 95% confidence intervals of the coloration changes at the off-center listening position during panning for different amplitude panning methods and loudspeaker apertures.

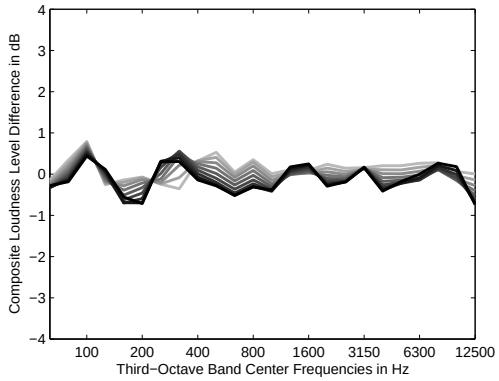
Figure 56 presents the medians and the corresponding confidence intervals for the 8 conditions evaluated at the off-center listening positions. Within the group with a loudspeaker aperture of $\Delta\phi_L = 45°$, Ambi$^{\mathrm{rV}}$ yields significantly more coloration than all other methods ($p < 0.001$). VBAP causes significantly more coloration than MDAP and Ambi$^{\mathrm{rE}}$ ($p < 0.018$). The difference between MDAP and Ambi$^{\mathrm{rE}}$ is not significant ($p = 0.80$).
Within the group with a loudspeaker aperture of $\Delta\phi_L = 22.5°$, all differences are significant ($p < 0.001$), except for the difference between MDAP and Ambi$^{\mathrm{rE}}$ ($p = 0.14$). Ambi$^{\mathrm{rV}}$ causes the most intense coloration changes, followed by VBAP. The least changes are perceived when using MDAP or Ambi$^{\mathrm{rE}}$.
Comparing both groups with different apertures, the larger aperture produces significantly less coloration ($p < 0.001$). In detail, the increase of the coloration with the increasing number of loudspeakers is significant for all panning methods ($p < 0.02$).



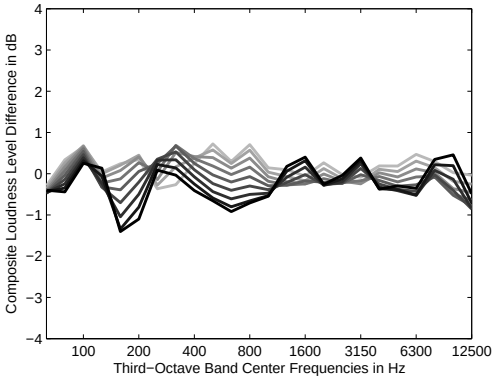
(a) VBAP with 8 loudspeakers



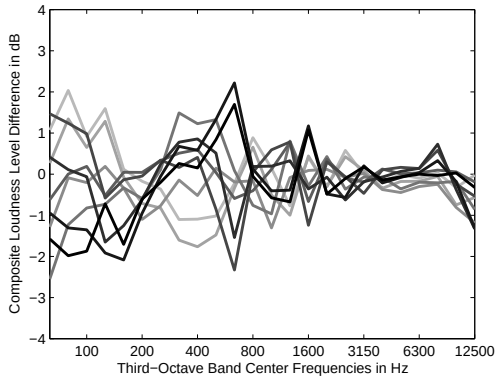
(b) MDAP with 8 loudspeakers



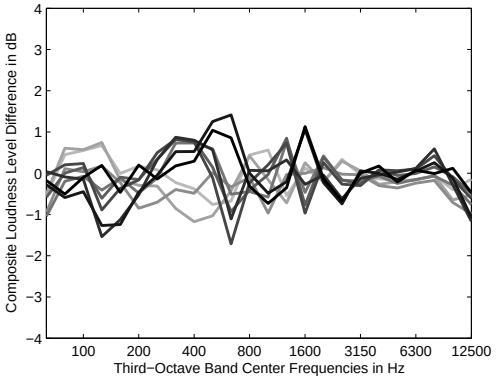
(c) Ambi$^{\mathrm{rE}}$ with 8 loudspeakers



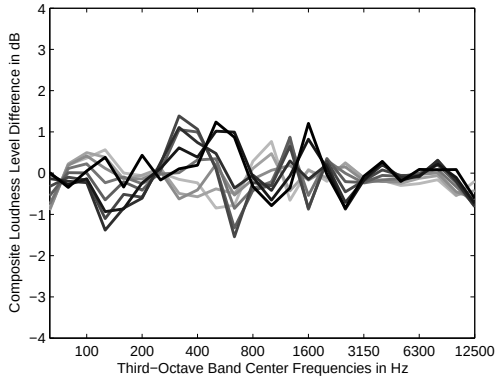
(d) Ambi$^{\mathrm{rV}}$ with 8 loudspeakers



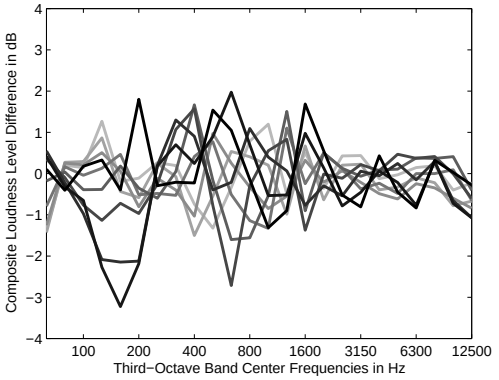
(e) VBAP with 16 loudspeakers



(f) MDAP with 16 loudspeakers



(g) Ambi$^{\mathrm{rE}}$ with 16 loudspeakers



(h) Ambi$^{\mathrm{rV}}$ with 16 loudspeakers

Figure 57: Incremental differences in the Composite Loudness Level Spectra for panning from 0° to 45° in 5° steps at the off-center listening position.

Figures 57(a) to 57(h) show the incremental changes in the CLL spectra for all conditions. The maximum values of the maximum spectral changes for each incremental step correlate with 0.71 to the median experimental results. All subjects reported to perceive a strong second auditory event that was split from the intended phantom source for both Ambi^{rv} conditions. This observation agrees with the localization results for the off-center listening position in Section 3.3.1. This splitting was also included in their assessment of the coloration changes. If the two Ambi^{rv} are omitted, the correlation between the median experimental results and the median values of the maximum of each incremental step increases to 0.97. The results for the off-center listening position differed from the results for the central listening position, yielding a correlation of only 0.53 to the results from the central listening position. However, the increase of the number of loudspeakers increased coloration. Also MDAP and Ambi^{rE} still yield least coloration.

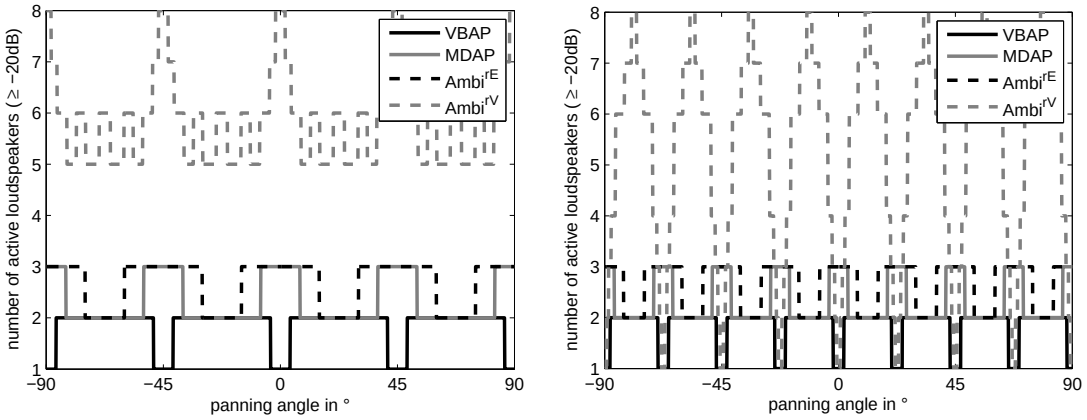
The theory of Theile [The80], that VBAP does not cause coloration strongly contradicts the experimental results from this section. Evidently, VBAP yields most coloration at the central listening position. Neither VBAP nor Ambisonics in their basic form (Ambi^{rv}) are suitable, when the least possible coloration is desired for all listening positions. MDAP and Ambi^{rE} are further developments of VBAP and Ambisonics, and they reduce the coloration.

5.4 Towards Simpler Predictors for Coloration of Moving Phantom Sources

The perceived coloration in the previous experiment could be explained by the sensitivity of fluctuation strength that is depending on the modulation frequency and the modulation depth. This section presents simpler predictors that set the stage for prediction of phantom source coloration at the central listening position without measurements but solely based on loudspeaker positions and gains.

5.4.1 Fluctuation in the Number of Active Loudspeakers

The strong coloration changes while panning a phantom source using VBAP could be attributed to the change in the number of active loudspeakers. There is a comb filter when two loudspeakers are active and there no comb filter when a single loudspeaker is active. It seems reasonable that the temporal fluctuation of the number of active loudspeakers is related to the coloration changes.



(a) panning with 8 loudspeakers, $\Delta\phi_L = 45°$ (b) panning with 16 loudspeakers, $\Delta\phi_L = 22.5°$

Figure 58: Number of active loudspeakers (all loudspeakers with gains $\geq$ -20dB of the loudest loudspeaker) during panning.

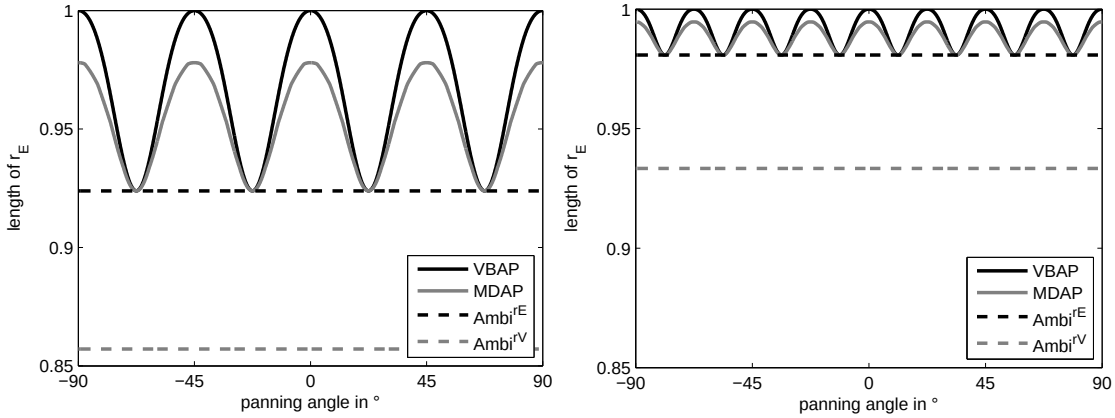
Figures 58(a) and 58(b) show the number of active loudspeakers in dependence of the panning angle using VBAP, MDAP, Ambi$^{\mathrm{rE}}$, and Ambi$^{\mathrm{rV}}$, as in the experiment. A loudspeaker is interpreted as active if its gain is $\geq$ -20dB of the gain of the loudest loudspeaker. The limit of -20dB is chosen, because a comb filter is not audible when the second impulse has a level of less than -20dB compared to the first impulse [BMW07]. VBAP mostly uses two loudspeakers, but for panning angles close to a loudspeaker, it only uses one loudspeaker. MDAP and Ambi$^{\mathrm{rE}}$ use two or three loudspeakers. For all three panning methods, this is not depending on the number of loudspeakers. The number of loudspeakers solely controls the speed to the fluctuation. Using Ambi$^{\mathrm{rV}}$, the number of active loudspeakers depend on the total number of loudspeakers and the order. For a larger number of loudspeakers

and a higher Ambisonics order, this results in a single active loudspeaker for panning on a loudspeaker position. The fluctuation between one and more active loudspeakers yields stronger comb filters as the fluctuation between two and three active loudspeakers, cf. Section 5.2.

The proposed predictor counts how often the number of active loudspeakers is changing during panning. The changes are divided into classes corresponding to the numbers of loudspeakers between which the changes occur. For instance, in the 8-loudspeaker case, cf. Figures 58(a), VBAP changes 16 times between 1 and 2 active loudspeakers, whereas MDAP and Ambi$^{\mathrm{r_E}}$ change 16 times between 2/3 loudspeakers. Ambi$^{\mathrm{r_V}}$ changes 64 times between 5/6, 16 times between 6/7, and 16 times between 7/8 loudspeakers. As mentioned before, a change between 1/2 loudspeakers yields a stronger coloration change than a change between 2/3 loudspeakers. Thus, the number of entries in each class is weighted according to its contribution to coloration changes. The weighting for each class c_{num} is chosen as $1/c_{num}^2$, i.e., 1 for the first class (1/2 loudspeakers), 1/4 for the second (2/3 loudspeakers), and so on. Summing up all weighted class entries yields a single value that incorporates both the modulation frequency and depth. The prediction for the 8 conditions correlate with 0.84 to the experimental results at the central listening position. As long as the threshold for activity stays between 17dB and 23dB, the correlation coefficient does not change.

5.4.2 Fluctuation in the Length of the Energy Vector

Another option to predict coloration changes during panning without measurements is the fluctuation of the energy vector length. Figures 59(a) and 59(b) show the length of the energy vector in dependence of the panning angle using VBAP, MDAP, Ambi$^{\mathrm{r_E}}$, and Ambi$^{\mathrm{r_V}}$.



(a) panning with 8 loudspeakers, $\Delta\phi_L = 45°$ (b) panning with 16 loudspeakers, $\Delta\phi_L = 22.5°$

Figure 59: Length of energy vector during panning.

Obviously, the length of the energy vector is independent of the panning angle for all Ambisonics conditions. For all panning methods, the lower limit of the energy vector length increases with the number of loudspeakers and the order. The length of 1 for VBAP panned on a loudspeaker indicates that only one loudspeaker is active. MDAP never reaches this value and yields a smaller amplitude in the fluctuation. Ambi$^{\mathrm{rE}}$ yields no fluctuation at all. Thus it seems that the fluctuation of the energy vector length is related to the coloration changes. However, this is not the case for Ambi$^{\mathrm{rV}}$, as it has no fluctuation but strong coloration changes.
As Ambi$^{\mathrm{rE}}$ created least coloration in the listening experiment, its energy vector is assumed be optimal. This optimal value $||\boldsymbol{r}_{\mathrm{E}}^{\mathrm{opt.}}|| = \cos\left({}^{\Delta\phi_L}/_{2}\right)$ corresponds to a loudspeaker pair with an aperture angle of $\Delta\phi_L$ and the equal gains. The deviation from $||\boldsymbol{r}_{\mathrm{E}}^{\mathrm{opt.}}||$ during panning indicates increasing coloration. Energy vectors that are larger than $||\boldsymbol{r}_{\mathrm{E}}^{\mathrm{opt.}}||$ tends towards the presence of single active loudspeakers. Coloration is increased if the maximum length is close to 1, e.g. for VBAP. Using the unit step function $u(x)$ which is 0 for $x < 0$ and 1 for $x \geq 0$, $u(\max(||\boldsymbol{r}_{\mathrm{E}}||) - 1)$ yields 1 for VBAP and 0 for all other panning methods. On the other hand, if the minimum length of the energy vector is smaller than $||\boldsymbol{r}_{\mathrm{E}}^{\mathrm{opt.}}||$ there are strong side lobes, e.g. for Ambi$^{\mathrm{rV}}$. In this case $u(\min(||\boldsymbol{r}_{\mathrm{E}}||) - ||\boldsymbol{r}_{\mathrm{E}}^{\mathrm{opt.}}||)$ yields 0, and otherwise 1. Linear regression yields coefficients to weight the above-mentioned deviations from $||\boldsymbol{r}_{\mathrm{E}}^{\mathrm{opt.}}||$ for a predictor of the coloration $C_{\boldsymbol{r}_{\mathrm{E}}}$ changes in the experiment

$$C_{\boldsymbol{r}_{\mathrm{E}}} = ||\boldsymbol{r}_{\mathrm{E}}^{\mathrm{opt.}}|| + a \cdot u(\max(||\boldsymbol{r}_{\mathrm{E}}||) - 1) + b \cdot u(\min(||\boldsymbol{r}_{\mathrm{E}}||) - ||\boldsymbol{r}_{\mathrm{E}}^{\mathrm{opt.}}||). \tag{25}$$

For $a = {}^{1}/_{16}$ and $b = -{}^{1}/_{23}$, the predicted $C_{\boldsymbol{r}_{\mathrm{E}}}$ values correlate to the median experimental results in the central listening position with 0.92. Nevertheless, the predictor is optimized for the experimental results and may require adaptation to other experiments.

5.5 Conclusion on Phantom Source Coloration

This chapter studied the coloration of phantom sources created by amplitude panning. It focused on changes of coloration as the panning-independent coloration can be equalized statically and is less annoying. The prediction of the coloration changes is based on the sum of the third-octave band levels at the ears, which is simplified version of the composite loudness level (CLL).

Results from the literature show that coloration changes of a single loudspeaker that is slowly rotated around a listener are hardly perceivable, although the spectral changes at each ear are obvious; a phenomenon called binaural decoloration. The results could be explained by the small incremental changes ($\leq$ 1dB) in the CLL spectrum. The CLL spectrum of a phantom source created by two loudspeakers exhibits a notch that was found to be depending on angle between the loudspeaker pair. The notch frequency and depth can be estimated from this angle. This holds true for three loudspeakers.

Applying the CLL to typical amplitude-panning scenarios showed that VBAP causes strong spectral differences for different panning angles. For MDAP and Ambi$^{\text{rE}}$, the spectrum is not flat, but its shape is rather panning-independent. Thus, it was expected that VBAP causes more coloration changes for phantom sources panned around the listener.

This hypothesis was examined by a listening experiment with moving phantom sources. At the central listening position, most coloration changes were perceived when using VBAP, followed by Ambi$^{\text{rV}}$. The coloration of MDAP and Ambi$^{\text{rE}}$ was similar and significantly less. For all panning methods, the use of more loudspeakers increased the perceived coloration changes as the modulation frequency of the spectral differences increases. A single value calculated from the spectral differences could predict the perceived coloration changes. The prediction is also possible without measurements, solely based on the positions and gains of the loudspeaker. One of these predictors is based on the fluctuation of the active loudspeaker count, another one examines the behavior of the energy vector during panning.

At an off-center listening position, the dependency on the number of loudspeakers is still present. MDAP and Ambi$^{\text{rE}}$ keep on causing the least coloration changes, and most coloration changes are perceived for Ambi$^{\text{rV}}$. This is dominated by the side lobes that produce a second auditory event, as already noticed in the localization experiment. The coloration at the off-center listening position can only be poorly predicted and would require modifications of the predictors, such as the incorporation of source splitting.

The results from this chapter contradict the theory of Theile [The80] that phantom sources created by VBAP were perceived without coloration. However, the presented results were evaluated for moving sources and not for static sources as in [The80]. In [Wit07], static phantom sources using VBAP were perceived similarly colored as a real sound source and less colored than wave field synthesis with a loudspeaker distance of > 3cm. As VBAP caused most coloration at the central listening position in this work, a comparison including moving sources also in wave field synthesis would be interesting.

Chapter VI

CONCLUSION

This work studied phantom sources created by amplitude-panning methods that use multiple loudspeakers in the horizontal plane, namely vector-base amplitude panning (VBAP), multiple-direction amplitude panning (MDAP), and Ambisonics without weighting (Ambi^{rV}) and with max $\boldsymbol{r}_{\text{E}}$ weighting (Ambi^{rE}). Besides the use of these distinct methods, generic studies using two and three loudspeakers with equal gain aim at the generalization of the results. The localization, width, and coloration of the phantom sources was evaluated by a series of listening experiments under typical non-anechoic studio listening conditions using pink noise as stimulus. In addition to the experimental results, predictive measures and models were presented.

Phantom Source Localization

As a starting point, localization experiments from the literature using pairwise panning on a standard $\pm 30°$ loudspeaker pair were compared to predictive models. The best localization prediction for broadband stimuli could be obtained by the direction of the energy vector. This vector solely depends on the positions and gains of the loudspeakers. Incorporating the directivity of human hearing as a direction-dependent weighting into the energy vector made it suitable for the prediction of lateral phantom sources. Own experiments at the central listening position revealed only small differences between the localization of VBAP, MDAP, Ambi^{rV}, and Ambi^{rE}. The weighted energy vector could also predict the results of this experiment and yielded a better accuracy than a state-of-the-art binaural localization model. Experimental results about localization at an off-center listening position showed that localization was dominated by the closest loudspeaker and yielded source splitting for Ambi^{rV}. However, all other panning methods achieved localization mismatch smaller than half the loudspeaker spacing.

Phantom Source Width

The investigation of phantom source width started with a listening experiment using symmetrically arranged loudspeaker pairs. It revealed the proportional relation between phantom source width and the loudspeaker aperture angle. Adding a loudspeaker in between the pair decreased the perceived width. The perception

could be well predicted by the length of the energy vector. The predictive accuracy for the $IACC_{E3}$ measure was fair, but the LF measure according to the ISO norm for room acoustics was not suitable at all. Nevertheless, the LF measure could be adapted to be a viable predictor of phantom source width by energetic superposition of the loudspeaker signals. Further experiments investigated the dependency of phantom source width on the panning angle for different panning methods. The largest dependency was found for VBAP as the number of active loudspeakers strongly depends on the panning angle. Once again, the length of the weighted energy vector could predict the experimental results reliably. This held true for experiments with lateral phantom sources. In this case, the assessment of phantom source width seemed to be more difficult.

Phantom Source Coloration

The studies about phantom source coloration focused on coloration changes during panning because they are most annoying. These changes were predicted by a binaural model that evaluates iterative differences in the sum of the third-octave band loudness levels (CLL) at both ears. The model could explain binaural decoloration of a single loudspeaker slowly rotating around a listeners. A frontal pair of loudspeakers with equal gain yielded a prominent notch in the CLL that could be estimated by the ITD that is caused by the loudspeaker positions. Comparing the CLL of different panning methods showed that VBAP exhibits the strongest dependency on the panning angle. This dependency could be proven by own listening experiments at the central listening position using a moving phantom source. The own experimental results contradict Theile's statement that coloration of phantom sources created by VBAP is imperceptible. The own experimental data could be predicted by the binaural model and simpler predictors that are based on the fluctuation of the number of active loudspeakers and the energy vector, respectively. At an off-center listening position, the source splitting of Ambi^{rV} was included in the coloration assessment and resulted in strong coloration changes.

Amplitude-Panning Methods

After the comparison of the different amplitude-panning methods, the question arises which one of them is the best, why and for what. VBAP yields the most distinct localization for all listening positions as it uses the smallest possible number of active loudspeakers. However, this causes a dependency of phantom source width on the panning angle and pronounced coloration changes during panning. Ambi^{rV} erroneously causes source splitting for off-center listening position and therefore strong coloration changes. Because Ambi^{rE} concentrates the energy into

the panning direction, it improves localization and lessens coloration. Similarly, MDAP successfully smooths the direction dependence of VBAP.
To summarize, accurate localization requires the smallest possible number of simultaneously active loudspeakers, whereas a constant width and little coloration requires the smallest possible fluctuation in the number of active loudspeakers. In the end, it depends on the application whether localization or smoothness in width and little coloration are preferred. Models, such as in [JDCZ10] that predict the overall quality of a spatial sound rendering system in a single value should consider preferences that depend on the application. It seems that MDAP and Ambi$^{\text{rE}}$ offer a good tradeoff between localization accuracy and smoothness. Both methods essentially are improvements of VBAP and Ambi$^{\text{rV}}$ thus some similarity in the distribution of the loudspeaker signals does not come as a surprise. In practice, the spread parameter of MDAP allows to be adjusted to the target application. Similarly, the AllRAD decoder [ZF12] can transform Ambi$^{\text{rE}}$ to become VBAP-like on demand [ZFP13].

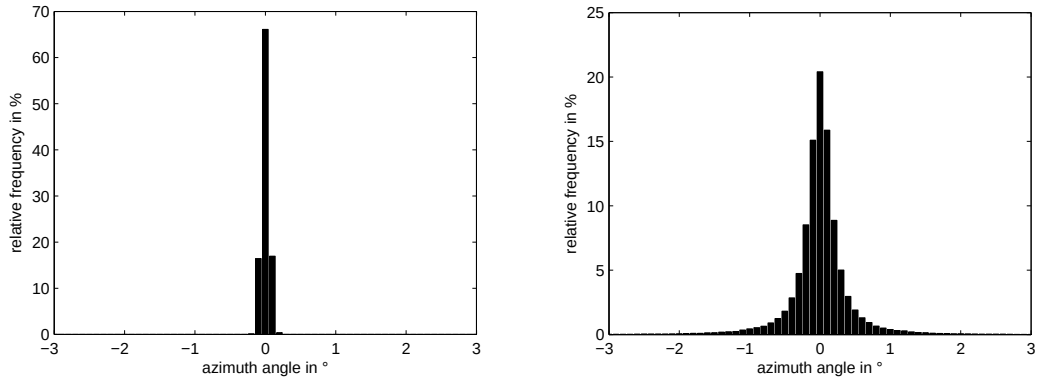
Predictors

The binaural measures and models based on summing localization predict phantom source localization, width, and coloration at the central listening position. They are suitable for all studied amplitude-panning methods independent of the number of active loudspeakers. This holds true for the energetic vector models, especially for the weighted energy vector that does not require measurements. Their predictive accuracy is comparable to the binaural models. This finding justifies the use of simple predictors that are solely based on loudspeaker positions and gains.
However, at the off-center listening position, the coloration was influenced by the source splitting, i.e. localization. It seems reasonable to develop models that predict the different parameters jointly, as it is suggested in the association model.

Appendix A

HEAD MOVEMENTS

In all experiments performed within the context of this thesis, the subjects were instructed not to move their head during the stimulus playback but to look at a certain direction, e.g. the direction of a loudspeaker. However, there was no mechanical head fixation in order to maintain a more natural listening situation. Nevertheless, the head movements of the subjects were recorded during the localization experiment, cf. Section 3.2, using the same optical tracking system as for the pointing device and they are analyzed in this section.
The analysis of the head movements requires their measurability, i.e., the measurement noise of the tracking system must be less than the head movements. Therefore the position and rotation angles have been measured for an immobile object for 10s. 94% of the measured positions (x, y, z) lie within the range of ±0.1mm and 98% of the measured rotation angles (azimuth, elevation, tilt) lie within the range of ±0.1°. Figure 60(a) exemplarily shows a histogram of the measurement noise for the azimuth angles.



(a) measurement noise of the tracking system (b) subjects' head movements

Figure 60: Histograms of the azimuth angles for the tracking system and the subjects' head movements.

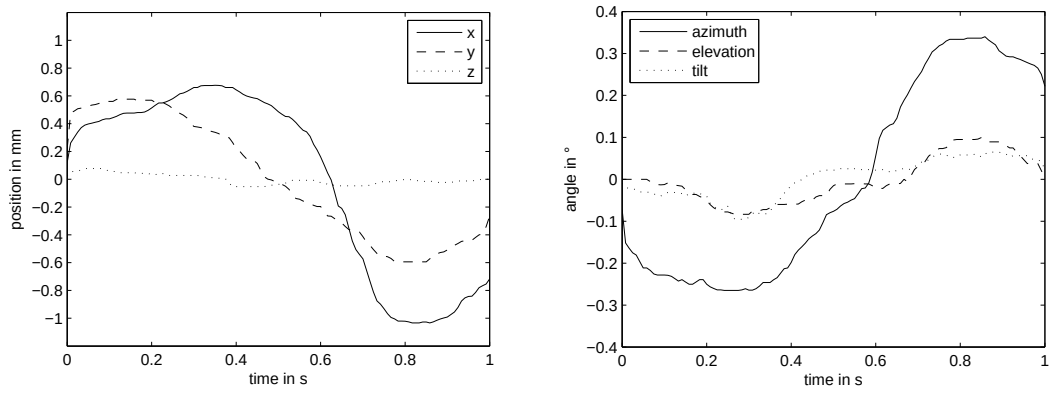
Figure 60(b) presents a histogram for the head movements in the azimuth. The histogram is based on 1s of tracking data for all 14 subjects that was recorded during the stimulus playback of each trial. For each trial and subject, the data was normalized to a mean value of 0. Summarizing, the data consists of 72 (trials) × 14 (subjects) × 120 (1s at a frame rate of 120Hz) = 120960 values. The results are similar to the amplitudes of head movements presented in [Bla83].

Table 17 compares the standard deviation of the measurement noise and the head movements for all 6 DOF. Except for the z position and the tilt angle, the measurement noise is less than one-tenth of the head movements. The largest position changes were recorded in the x and y directions. However, their standard deviations are smaller than 2mm. There is nearly no head movement in the z direction. The rotation is focused on azimuth and elevation and yields standard deviations below 1°.

standard deviation	**x** in mm	**y** in mm	**z** in mm	**azimuth** in °	**elevation** in °	**tilt** in °
measurement noise	0.07	0.07	0.06	0.05	0.04	0.07
head movements	1.84	1.53	0.48	0.55	0.75	0.34

Table 17: Standard deviations of the measurement noise and the head movements for all 6 DOF.

Although these small movements can not lead to a conscious improvement in the localization performance [Bla83], it is interesting that these movements do not have a random shape. Figures 61(a) and 61(b) show 1s of a typical head movement during stimulus playback. For all subjects the movements look similar and have a sinusoidal shape with a periodicity of approximately 1Hz. The periodicity is similar to the speed of conscious head movements described in [BKM07] where the subjects were allowed to move their heads freely, i.e., they were neither instructed to keep their heads still nor was there any mechanical head fixation. However, the amplitude of the conscious head movements was much bigger compared to the unconscious movements recorded here.



(a) head position over time

(b) head rotation over time

Figure 61: Exemplary head movement during 1 s of stimulus playback using a frame rate of 120 Hz and 20 taps median filter.

REFERENCES

[ANS06] ANSI, *ANSI/ASA S1.6-1984 (R2006) Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements*, American National Standards Institute Std., 2006.

[AS08] J. Ahrens and S. Spors, "An Analytical Approach to Sound Field Reproduction using Circular and Spherical Loudspeaker Distributions," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 988–999, December 2008.

[Bau61] B. B. Bauer, "Phasor analysis of some stereophonic phenomena," *The Journal of the Acoustical Society of America*, vol. 33, no. 11, pp. 1536–1539, 1961. [Online]. Available: http://link.aip.org/link/?JAS/33/1536/1

[BBE85] J. C. Bennett, K. Barker, and F. O. Edeko, "A New Approach to the Assessment of Stereophonic Sound System Performance," *J. Audio Eng. Soc*, vol. 33, no. 5, pp. 314–321, 1985. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=4449

[BDG+07] S. Bertet, J. Daniel, L. Gros, E. Parizet, and O. Warusfel, "Investigation of the Perceived Spatial Resolution of Higher Order Ambisonics Sound Fields: A Subjective Evaluation Involving Virtual and Real 3D Microphones," in *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*, 3 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13925

[BdVV93] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993. [Online]. Available: http://link.aip.org/link/?JAS/93/2764/1

[Ber88] A. J. Berkhout, "A Holographic Approach to Acoustic Control," *J. Audio Eng. Soc*, vol. 36, no. 12, pp. 977–995, 1988. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=5117

[BF11] S. Braun and M. Frank, "Localization of 3D Ambisonic Recordings and Ambisonic Virtual Sources," in *International Conference on Spatial Audio*, November 2011.

[BHL06] E. Benjamin, A. Heller, and R. Lee, "Localization in Horizontal-Only Ambisonic Systems," in *Audio Engineering Society Convention 121*, 10 2006. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13801

[BK04] M. Bouéri and C. Kyriakakis, "Audio signal decorrelation based on a critical band approach," in *Audio Engineering Society Convention 117*, 10 2004. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12948

[BKM07] T. Brookes, C. Kim, and R. Mason, "An investigation into head movements made when evaluating various attributes of sound," in *Audio Engineering Society Convention 122*, 5 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14016

[BL86a] J. Blauert and W. Lindemann, "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," *The Journal of the Acoustical Society of America*, vol. 79, no. 3, pp. 806–813, 1986. [Online]. Available: http://link.aip.org/link/?JAS/79/806/1

[BL86b] ——, "Auditory spaciousness: Some further psychoacoustic analyses," *The Journal of the Acoustical Society of America*, vol. 80, no. 2, pp. 533–542, 1986. [Online]. Available: http://link.aip.org/link/?JAS/80/533/1

[Bla83] J. Blauert, *Spatial Hearing*. MIT Press, 1983.

[Bla02] M. Blau, "Difference limens for measures of apparent source width," in *Forum Acusticum, Sevilla, Spain*, 2002.

[Blu58] A. D. Blumlein, "British patent specification 394,325 (improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems)," *J. Audio Eng. Soc*, vol. 6, no. 2, pp. 91–98, 130, 1958. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=233

[BM81] M. Barron and A. Marshall, "Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure," *Journal of Sound and Vibration*, vol. 77, no. 2, pp. 211 – 232, 1981. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022460X8180020X

[BM13] R. Baumgartner and P. Majdak, "Modeling Sound Localization of Amplitude-Panned Phantom Sources in Sagittal Planes," in *Audio Engineering Society Convention 134*, 5 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16810

[BMCQGAOB11] E. Blanco-Martin, F. J. Casajús-Quirós, J. J. Gómez-Alfageme, and L. I. Ortiz-Berenguer, "Objective measurement of sound event localization in horizontal and median planes," *Journal of the Audio Engineering Society*, vol. 59, no. 3, pp. 124–136, 2011.

[BMW07] S. Brunner, H.-J. Maempel, and S. Weinzierl, "On the audibility of comb filter distortions," in *Audio Engineering Society Convention 122*, 5 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14032

[BR99] J. Berg and F. Rumsey, "Spatial attribute identification and scaling by repertory grid technique and other methods," in *16th International Conference on Spatial Sound Reproduction, Audio Engineering Society*, 1999, pp. 51–66.

[Brü01] M. Brüggen, "Coloration and binaural decoloration in natural environments," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 400–406, 2001. [Online]. Available: http://www.ingentaconnect.com/content/dav/aaua/2001/00000087/00000003/art00012

[BS13] A. D. Brown and G. C. Stecker, "The precedence effect: Fusion and lateralization measures for headphone stimuli lateralized by interaural time and level differences," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 2883–2898, 2013. [Online]. Available: http://link.aip.org/link/?JAS/133/2883/1

[CB89] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc*, vol. 37, no. 1/2, pp. 3–19,

1989. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=6108

[CDL93] T. J. Cox, W. J. Davies, and Y. W. Lam, "The sensitivity of listeners to early sound field changes in auditoria," *Acta Acustica united with Acustica*, vol. 79, no. 1, pp. 27–41, 1993. [Online]. Available: http://www.ingentaconnect.com/content/dav/aaua/1993/00000079/00000001/art00007

[CDT+75] R. C. Cabot, D. A. Dorans, I. S. Tackel, D. B. Wilsone, and H. E. Breed, "Localization effects in the quadraphonic sound field," in *Audio Engineering Society Convention 52*, 10 1975. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=2347

[CG74] P. G. Craven and M. A. Gerzon, "Coincident microphone simulation covering three dimensional space and yielding various directional outputs," *British patent 1512514*, 1974.

[Cra03] P. G. Craven, "Continuous surround panning for 5-speaker reproduction," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, 6 2003.

[CS72] D. H. Cooper and T. Shiga, "Discrete-matrix multichannel stereo," *Journal of the Audio Engineering Society*, vol. 20, no. 5, pp. 346–360, 1972.

[CT03] D. Cabrera and S. Tilley, "Parameters for auditory display of height and size," in *Proc. ICAD*, 2003.

[CW07] S. Choisel and F. Wickelmaier, "Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference," *Journal of the Acoustical Society of America*, vol. 121, no. 1, pp. 388–400, 2007.

[Dan01] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scénes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Université Paris 6, 2001.

[Dan03] ——, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic

format," in *23rd International Conference, Audio Engineering Society*, Copenhagen, Denmark, May 2003.

[dVHB01] D. de Vries, E. M. Hulsebos, and J. Baan, "Spatial fluctuations in measures for spaciousness," *The Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 947–954, 2001. [Online]. Available: http://link.aip.org/link/?JAS/110/947/1

[Fel75] P. Fellget, "Ambisonics. Part one: General system description," *Studio Sound*, vol. 17, pp. 20–22, 1975.

[FMS11] M. Frank, G. Marentakis, and A. Sontacchi, "A simple technical measure for the perceived source width," in *Fortschritte der Akustik, DAGA*, Düsseldorf, 2011.

[FMSZ10] M. Frank, L. Mohr, A. Sontacchi, and F. Zotter, "Flexible and Intuitive Pointing Method for 3-D Auditory Localization Experiments," in *Audio Engineering Society Conference: 38th International Conference: Sound Quality Evaluation*, 6 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15469

[FNCS08] F. Fazi, P. Nelson, J. Christensen, and J. Seo, "Surround system based on three dimensional sound field reconstruction," in *125th Convention of the Audio Engineering Society*, San Fransisco, USA, 2008.

[Fra09] M. Frank, "Perzeptiver Vergleich von Schallfeldreproduktionsverfahren unterschiedlicher räumlicher Bandbreite," Master's thesis, University of Music and Performing Arts, Institute of Electronic Music and Acoustics, Graz, 2009.

[Fra13] ——, "Source width of frontal phantom sources: Perception, measurement, and modeling," *Archives of Acoustics*, vol. 38, 2013, in press.

[FS12] M. Frank and A. Sontacchi, "Performance review of an expert listening panel," in *Fortschritte der Akustik, DAGA*, Darmstadt, 2012.

[FSH10] M. Frank, A. Sontacchi, and R. Höldrich, "Training and guidance tool for listening panels," in *Fortschritte der Akustik, DAGA*, Berlin, 2010.

[FSLO12] M. Frank, A. Sontacchi, T. Lindenbauer, and M. Opitz, "Subjective sound quality evaluation of a codec for digital wireless transmission," in *Audio Engineering Society Convention 132*, 4 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16224

[FU99] A. Farina and E. Ugolotti, "Subjective Comparison Between Stereo Dipole and 3D Ambisonic Surround Systems for Automotive Applications," in *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*, 3 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8006

[FZ08] M. Frank and F. Zotter, "Localization experiments using different 2D Ambisonics decoders," in *25. Tonmeistertagung, Leipzig*, 2008.

[FZ13] ——, "Simple technical prediction of phantom source widening," in *Fortschritte der Akustik, AIA-DAGA*, Meran, 2013.

[Ger73] M. A. Gerzon, "With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, pp. 2–10, 1973.

[Ger75] ——, "Ambisonics. Part two: Studio techniques," *Studio Sound*, vol. 17, pp. 24–26, 1975.

[Ger92] ——, "General metatheory of auditory localisation," in *Audio Engineering Society Convention 92*, 3 1992. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=6827

[Ger93] ——, "Stereophonic signal processor generating pseudo stereo signals," *Patent, WO 93/25055*, 1993.

[GK04] C. Guastavino and B. F. G. Katz, "Perceptual evaluation of multi-dimensional spatial audio reproduction," *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 1105–1115, 2004. [Online]. Available: http://link.aip.org/link/?JAS/116/1105/1

[Haa72] H. Haas, "The influence of a single echo on the audibility of speech," *J. Audio Eng. Soc*, vol. 20, no. 2, pp. 146–159, 1972. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=2093

[Ham05] K. O. R. Hamasaki, Kimio; Hiyama, "The 22.2 Multichannel Sound System and Its Application," in *Audio Engineering Society Convention 118*, 5 2005. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13122

[Har83] W. M. Hartmann, "Localization of sound in rooms," *The Journal of the Acoustical Society of America*, vol. 74, no. 5, pp. 1380–1391, 1983. [Online]. Available: http://link.aip.org/link/?JAS/74/1380/1

[HBO95] T. Hidaka, L. L. Beranek, and T. Okano, "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," *The Journal of the Acoustical Society of America*, vol. 98, no. 2, pp. 988–1007, 1995. [Online]. Available: http://link.aip.org/link/?JAS/98/988/1

[Hes07] W. Hess, "Time-variant binaural activity characteristics as indicator of auditory spatial attributes," Ph.D. dissertation, Ruhr-Universität Bochum, 2007.

[Hey86] R. C. Heyser, "Instantaneous intensity," in *Audio Engineering Society Convention 81*, 11 1986. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=5031

[Hir08] V. Hirvonen, Toni; Pulkki, "Perceived spatial distribution and width of horizontal ensemble of independent noise signals as function of waveform and sample length," in *Audio Engineering Society Convention 124*, 5 2008. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14538

[HOT+07] J. Huang, J. Ootomo, K. Tanno, W. Hatano, and A. Saji, "The precedence effect of sound from side direction," in *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*, 3 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13942

[HP06] T. Hirvonen and V. Pulkki, "Perception and analysis of selected auditory events with frequency-dependent directions," *J. Audio Eng. Soc*, vol. 54, no. 9, pp. 803–814, 2006. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13902

[Iid08] K. Iida, "Estimation of sound source elevation by extracting the vertical localization cues from binaural signals," in *Proceedings of Meetings on Acoustics*, vol. 4, 2008, p. 050002.

[ISO09] ISO, *ISO 3382-1:2009: Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces*, International Organization for Standardization. Std., 2009.

[ITU97] ITU, *ITU-R BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, International Telecommunication Union Std., 1997.

[ITU06] ——, *ITU-R BS.775-2: Multichannel stereophonic sound system with and without accompanying picture*, International Telecommunication Union Std., 2006.

[Jah63] G. Jahn, "Zum Unterschied zwischen einohrigem und zweiohrigem Hören," *Hochfrequenztechnik und Elektroakustik*, vol. 72, 1963.

[JDCZ10] P. Jackson, M. Dewhirst, R. Conetta, and S. Zielinski, "Estimates of perceived spatial quality across the listening area," in *Audio Engineering Society Conference: 38th International Conference: Sound Quality Evaluation*, 6 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15478

[Jef48] L. A. Jeffress, "A place theory of sound localization," *Journal of comparative and physiological psychology*, vol. 41, no. 1, pp. 35–39, 1948.

[Jes73] M. Jessel, *Acoustique théorique: propagation et holophonie*. Paris: Masson et Cie, 1973.

[JV59] G. Jahn and S. Vogelsang, "Die einohrige Richtcharakteristik des menschlichen Gehörs," *Hochfrequenztechnik und Elektroakustik*, vol. 68, 1959.

[KA12] T. Kimura and H. Ando, "Listening test for three-dimensional audio system based on multiple vertical panning," in *Acoustics 2012*, Hong Kong, 2012.

[KBBF07] G. Kearney, E. Bates, F. Boland, and D. Furlong, "A comparative study of the performance of spatialization techniques for a distributed audience in a concert hall environment," in *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*, 6 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13948

[Ken95] G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music Journal*, vol. 19, no. 4, pp. 71–87, 1995.

[KP11] M. J. Kin and P. Plaskota, "Comparison of sound attributes of multichannel and mixed-down stereo recordings," *Archives of Acoustics*, vol. 36, no. 2, pp. 333–345, May 2011.

[KPJH99] M. Karjalainen, E. Piirilä, A. Järvinen, and J. Huopaniemi, "Comparison of Loudspeaker Equalization Methods Based on DSP Techniques," *J. Audio Eng. Soc*, vol. 47, no. 1/2, pp. 14–31, 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12117

[LB02] E. H. A. Langendijk and A. W. Bronkhorst, "Contribution of spectral cues to human sound localization," *The Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1583–1596, 2002. [Online]. Available: http://link.aip.org/link/?JAS/112/1583/1

[Lea59] 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. [Online]. Available: http://link.aip.org/link/?JAS/31/977/1

[Lew98] J. Lewald, "The effect of gaze eccentricity on perceived sound direction and its relation to visual localization," *Hearing Research*, vol. 115, no. 1–2, pp. 206–216, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378595597001901

[Lin86a] 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. [Online]. Available: http://link.aip.org/link/?JAS/80/1608/1

[Lin86b] ——, "Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front," *The Journal of the Acoustical Society of America*, vol. 80, no. 6, pp. 1623–1630, 1986. [Online]. Available: http://link.aip.org/link/?JAS/80/1623/1

[LPLP12] M.-V. Laitinen, T. Pihlajamäki, S. Lösler, and V. Pulkki, "Influence of resolution of head tracking in synthesis of binaural audio," in *Audio Engineering Society Convention 132*, 4 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16261

[Mac08] P. Mackensen, "Auditive localization & head movements," Ph.D. dissertation, TU Berlin, 2008.

[Mak62] Y. Makita, "On the directional localization of sound in the stereophonic sound field," EBU, Review Part A, 73, Tech. Rep., 1962.

[ME02] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. 2, Orlando, 13th May 2002, pp. 1781–1784. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1006109

[Mer07] J. Merimaa, "Energetic sound field analysis of stereo and multichannel loudspeaker reproduction," in *Audio Engineering Society Convention 123*, 10 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14315

[MFRB01] R. Mason, N. Ford, F. Rumsey, and B. D. Bruyn, "Verbal and nonverbal elicitation techniques in the subjective assessment of spatial sound reproduction," *J. Audio Eng. Soc*, vol. 49, no. 5, pp. 366–384, 2001. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10190

[MI05] M. Morimoto and K. Iida, "Appropriate frequency bandwidth in measuring interaural cross-correlation as a physical measure of auditory source width," *Acoustical Science and Technology*, vol. 26, no. 2, pp. 179–184, 2005.

[MLGM08] P. Majdak, B. Laback, M. Goupell, and M. Mihocic, "The accuracy of localizing virtual sound sources: Effects of pointing method and visual environment," in *Audio Engineering Society Convention 124*, 5 2008. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14537

[MM88] M. Morimoto and Z. Maekawa, "Effects of low frequency components on auditory spaciousness," *Acta Acustica united with Acustica*, vol. 66, no. 4, pp. 190–196, 1988. [Online]. Available: http://www.ingentaconnect.com/content/dav/aaua/1988/00000066/00000004/art00004

[MM95] D. G. Malham and A. Myatt, "3D Sound Spatialization using Ambisonic Techniques," *Computer Music Journal*, vol. 19, no. 4, pp. 58–70, 1995.

[MM02] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2219–2236, 2002. [Online]. Available: http://link.aip.org/link/?JAS/111/2219/1

[MPG90] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "Auditory filter shapes at low center frequencies," *The Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 132–140, 1990. [Online]. Available: http://link.aip.org/link/?JAS/88/132/1

[MRZ08] P. Marins, F. Rumsey, and S. Zielinski, "Sound quality assessment: Cardinal concepts," in *Proceedings of the Institute of Acoustics*, vol. 30, no. 6, 2008, pp. 297–305.

[MW08] D. Moore and J. Wakefield, "The design of ambisonic decoders for the itu 5.1 layout with even performance characteristics," in *Audio Engineering Society Convention 124*, 5 2008. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14603

[MW10] ——, "A design tool to produce optimized ambisonic decoders," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, 10 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15541

[MWCQ99a] G. Martin, W. Woszczyk, J. Corey, and R. Quesnel, "Sound source localization in a five-channel surround sound reproduction system," in *Audio Engineering Society Convention 107*, 9 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8186

[MWCQ99b] ——, "Controlling phantom image focus in a multichannel reproduction system," in *Audio Engineering Society Convention 107*, 9 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8184

[NCW12] M. Noisternig, T. Carpentier, and O. Warusfel, "ESPRO 2.0 - Implementation of a surrounding 350-loudspeaker array for 3D sound field reproduction," in *25th AES UK Conference: Spatial Audio in Today's 3D World in association with the 4th International Symposium on Ambisonics and Spherical Acoustics*, 2012.

[Oka02] T. Okano, "Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections," *The Journal of the Acoustical Society of America*, vol. 111, no. 1, pp. 217–229, 2002. [Online]. Available: http://link.aip.org/link/?JAS/111/217/1

[OPK01] K. Ono, V. Pulkki, and M. Karjalainen, "Binaural modeling of multiple sound source perception: Methodology and coloration experiments," in *Audio Engineering Society Convention 111*, 11 2001. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=9884

[OPK02] ——, "Binaural modeling of multiple sound source perception: Coloration of wideband sound," in *Audio Engineering Society Convention 112*, 4 2002. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=11331

[Orb70] R. Orban, "Further Thoughts on "A Rational Technique for Synthesizing Pseudo-Stereo from Monophonic Sources"," *J. Audio Eng. Soc*, vol. 18, no. 4, pp. 443– 444, 1970. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=1484

[PB82] D. R. Perrott and T. N. Buell, "Judgments of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise," *The Journal of the Acoustical Society of America*, vol. 72, no. 5, pp. 1413–1417, 1982. [Online]. Available: http://link.aip.org/link/?JAS/72/1413/1

[PK01] V. Pulkki and M. Karjalainen, "Localization of amplitude-panned virtual sources I: Stereophonic panning," *J. Audio Eng. Soc*, vol. 49, no. 9, pp. 739–752, 2001. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10180

[PKH99] V. Pulkki, M. Karjalainen, and J. Huopaniemi, "Analyzing virtual sound source attributes using a binaural auditory model," *J. Audio Eng. Soc*, vol. 47, no. 4, pp. 203–217, 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12110

[Pol96] M. A. Poletti, "The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems," *J. Audio Eng. Soc*, vol. 44, no. 11, pp. 948–963, 1996. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7879

[Pol00] ——, "A Unified Theory of Horizontal Holographic Sound Systems," *J. Audio Eng. Soc*, vol. 48, no. 12, pp. 1155–1182, 2000. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12033

[Pot06] G. Potard, "3D-audio object oriented coding," Ph.D. dissertation, University of Wollongong, 2006.

[PR11] Y. L. Parodi and P. Rubak, "Analysis of design parameters for crosstalk cancellation filters applied to different loudspeaker configurations," *J. Audio Eng. Soc*, vol. 59, no. 5, pp. 304–320, 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15931

[Pul97] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc*, vol. 45, no. 6, pp. 456–466, 1997. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7853

[Pul99] ——, "Uniform spreading of amplitude panned virtual sources," in *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, 1999, pp. 187–190.

[Pul01a] ——, "Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning," *J. Audio Eng. Soc*, vol. 49, no. 9, pp. 753–767, 2001. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10179

[Pul01b] ——, "Spatial sound generation and perception by amplitude panning techniques," Ph.D. dissertation, Helsinki University of Technology, 2001.

[Pul02] ——, "Compensating displacement of amplitude-panned virtual sources," in *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, 6 2002. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=11138

[Pul07] ——, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc*, vol. 55, no. 6, pp. 503–516, 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14170

[RS87] A. Ramsteiner and G. Spikofski, "Ermittlung von Wahrnehmbarkeitsschwellen für Klangfarbenunterschiede unter Verwendung eines diffusfeldentzerrten Kopfhörers," *Fortschritte der Akustik, DAGA*, 1987.

[Rum02] F. Rumsey, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm," *J. Audio Eng. Soc*, vol. 50, no. 9, pp. 651–666, 2002. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=11067

[RZKB05] F. Rumsey, S. Zieliński, R. Kassier, and S. Bech, "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," *The*

Journal of the Acoustical Society of America, vol. 118, no. 2, pp. 968–976, 2005. [Online]. Available: http://link.aip.org/link/?JAS/118/968/1

[Sal95] A. M. Salomons, "Coloration and binaural decoloration of sound due to reflections," Ph.D. dissertation, Technische Universiteit Delft, 1995.

[Sch58] M. R. Schröder, "An artificial stereophonic effect obtained from a single audio signal," *J. Audio Eng. Soc*, vol. 6, no. 2, pp. 74–79, 1958. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=238

[SE06] V. P. Sivonen and W. Ellermeier, "Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2965–2980, 2006. [Online]. Available: http://link.aip.org/link/?JAS/119/2965/1

[See03] B. Seeber, "Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode," Ph.D. dissertation, Technischen Universität München, 2003.

[Sen12] E. Sengpiel. (2012, May) Gleichungen für die Pegeldifferenz- und Laufzeitdifferenz-Lokalisationskurve. Accessed on June 2012. [Online]. Available: http://www.sengpielaudio.com/GleichungenDLundDt.pdf

[SGS74] M. R. Schröder, D. Gottlob, and K. F. Siebrasse, "Comparative study of european concert halls: correlation of subjective preference with geometric and acoustic parameters," *The Journal of the Acoustical Society of America*, vol. 56, no. 4, pp. 1195–1201, 1974. [Online]. Available: http://link.aip.org/link/?JAS/56/1195/1

[SM10] L. S. R. Simon and R. Mason, "Time and level localization curves for a regularly-spaced octagon loudspeaker array," in *Audio Engineering Society Convention 128*, 5 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15376

[SMR09] L. S. R. Simon, R. Mason, and F. Rumsey, "Localization curves for a regularly-spaced octagon loudspeaker array," in

Audio Engineering Society Convention 127, 10 2009. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15110

[Sno53] W. B. Snow, "Basic principles of stereophonic sound," *Jour. SMPTE*, vol. 61, pp. 567–587, 1953.

[SP11] O. Santala and V. Pulkki, "Directional perception of distributed sound sources," *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1522–1530, 2011. [Online]. Available: http://link.aip.org/link/?JAS/129/1522/1

[SPH09] A. Sontacchi, H. Pomberger, and R. Höldrich, "Recruiting and evaluation process of an expert listening panel," in *NAG/DAGA International Conference on Acoustics*, Rotterdam, 2009.

[SRA08] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *Audio Engineering Society Convention 124*, 5 2008. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14488

[SS34] J. C. Steinberg and W. B. Snow, "Symposium on wire transmission of symphonic music and its reproduction in auditory perspective: Physical factors," *Bell Systems Technical Journal*, vol. XIII, no. 2, April 1934.

[SSJW10] R. Schleicher, S. Spors, D. Jahn, and R. Walter, "Gaze as a measure of sound source localization," in *Audio Engineering Society Conference: 38th International Conference: Sound Quality Evaluation*, 6 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15466

[Str07] J. W. Strutt, "On our perception of sound direction," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 13, no. 74, pp. 214–232, 1907.

[SVP+09] O. Santala, H. Vertanen, J. Pekonen, J. Oksanen, and V. Pulkki, "Effect of listening room on audio quality in ambisonics reproduction," in *Audio Engineering Society Convention 126*, 5 2009. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14860

[The80] G. Theile, "Über die Lokalisation im überlagerten Schallfeld," Ph.D. dissertation, Technischen Universität Berlin, 1980.

[The91] ——, "On the naturalness of two-channel stereo sound," *J. Audio Eng. Soc*, vol. 39, no. 10, pp. 761–767, 1991. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=5963

[Thu94] L. L. Thurstone, "A law of comparative judgment," *Psychological Review*, vol. 101, no. 2, pp. 266–270, 1994.

[TP77] G. Theile and G. Plenge, "Localization of lateral phantom sources," *J. Audio Eng. Soc*, vol. 25, no. 4, pp. 196–200, 1977. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=3376

[TW11] G. Theile and H. Wittek, "Principles in surround recordings with height," in *Audio Engineering Society Convention 130*, 5 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15870

[vDS11] J. van Dorp Schuitman, "Auditory modelling for assessing room acoustics," Ph.D. dissertation, Technische Universiteit Delft, 2011.

[Völ11] F. Völk, "System theory of binaural synthesis," in *Audio Engineering Society Convention 131*, 10 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16093

[WA09] Y. J. Wu and T. D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 1, pp. 107–116, 2009.

[Wen63] K. Wendt, "Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie," Ph.D. dissertation, RWTH Aachen, 1963.

[WFZ13] F. Wendt, M. Frank, and F. Zotter, "Application of localization models for vertical phantom sources," in *Fortschritte der Akustik, AIA-DAGA*, Meran, 2013.

[Wit07] H. Wittek, "Perceptual differences between wavefield synthesis and stereophony," Ph.D. dissertation, University of Surrey, 2007.

[WK92] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648–1661, 1992. [Online]. Available: http://link.aip.org/link/?JAS/91/1648/1

[WK97] ——, "Monaural sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1050–1063, 1997. [Online]. Available: http://link.aip.org/link/?JAS/101/1050/1

[WNR73] H. Wallach, E. B. Newman, and M. R. Rosenzweig, "The precedence effect in sound localization (tutorial reprint)," *J. Audio Eng. Soc*, vol. 21, no. 10, pp. 817–826, 1973. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10299

[WRGS13] H. Wierstorf, A. Raake, M. Geier, and S. Spors, "Perception of focused sources in wave field synthesis," *J. Audio Eng. Soc*, vol. 61, no. 1/2, pp. 5–16, 2013. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16663

[WRS12] H. Wierstorf, A. Raake, and S. Spors, "Localization of a virtual point source within the listening area for wave field synthesis," in *Audio Engineering Society Convention 133*, 10 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16485

[WS12] H. Wierstorf and S. Spors, "Sound field synthesis toolbox," in *Audio Engineering Society Convention 132*, 4 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16605

[WT02] H. Wittek and G. Theile, "The recording angle - based on localisation curves," in *Audio Engineering Society Convention 112*, 4 2002. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=11307

[Zah02] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1832–1846, 2002. [Online]. Available: http://link.aip.org/link/?JAS/111/1832/1

[ZF90] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Springer-Verlag, Heidelberg, 1990.

[ZF12] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *J. Audio Eng. Soc*, vol. 60, no. 10, pp. 807–820, 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16554

[ZF13] ——, "Efficient phantom source widening," *Archives of Acoustics*, vol. 38, no. 1, pp. 27–37, 2013.

[ZFMS11] F. Zotter, M. Frank, G. Marentakis, and A. Sontacchi, "Phantom source widening with deterministic frequency dependent time delays," in *DAFx-11*, 2011.

[ZFP13] F. Zotter, M. Frank, and H. Pomberger, "Comparison of energy-preserving and all-round ambisonic decoders," in *Fortschritte der Akustik, AIA-DAGA*, Meran, 2013.

[ZFS10] F. Zotter, M. Frank, and A. Sontacchi, "The Virtual t-design Ambisonics-Rig using VBAP," in *Proc. EAA EUROREGIO*, Ljubljana, 2010.

[ZK01] N. Zacharov and K. Koivuniemi, "Audio descriptive analysis & mapping of spatial sound displays," in *Proceedings of the 2001 International Conference on Auditory Display*, 2001.

[ZPN12] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-Preserving Ambisonic Decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012. [Online]. Available: http://www.ingentaconnect.com/content/dav/aaua/2012/00000098/00000001/art00004