Source Width in Music Production. Methods in Stereo, Ambisonics, and Wave Field Synthesis

Tim Ziemer

Abstract Source width of musical instruments, measured in degrees, is a matter of source extent and the distance of the observer. In contrast to that, perceived source width is a matter of psychological organization of sound. It is influenced by the sound radiation characteristics of the source and by the room acoustics and restricted by masking and by localization accuracy. In this chapter perceived source width in psychoacoustics and apparent source width in room acoustical research are revisited. Source width in music recording and production practice in stereo and surround as well as in ambisonics and wave field synthesis are addressed. After the review of the literature an investigation is introduced. The radiation characteristics of 10 musical instruments are measured at 128 angles and the radiated sound is propagated to potential listening positions at 3 different distances. Here, monaural and binaural sound quantities are calculated. By means of multiple linear regression, the physical source extent is predicted by sound field quantities. The combination of weighted interaural phase differences in the sensitive frequency region together with the number of partials in the quasi-stationary part of instrumental sounds shows significant correlation with the actual source extent of musical instruments. The results indicate that these parameters might have a relevant effect on perceived source extent as well. Consequently, acoustic control over these parameters will increase psychoacoustic control concerning perceived source extent in audio systems.

1 Introduction

Due to extensive and well-elaborated investigations in the field of psychoacoustics and subjective room acoustics within the last hundred years, a lot of knowledge about the auditory perception of source extent has been acquired. It is outlined in

T. Ziemer (✉)
Institute of Systematic Musicology, University of Hamburg,
Neue Rabenstr. 13, 20354 Hamburg, Germany
e-mail: tim.ziemer@uni-hamburg.de

© Springer International Publishing AG 2017

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_10

the following section. In music recording, mixing and mastering practice, several methods to control the perceived source extent have been established for channel based audio systems like stereo and surround. More recently, novel approaches for object based audio systems like ambisonics and wave field synthesis have been proposed. These are revisited and examined from a psychoacoustic point of view. Following this theoretic background, an investigation to illuminate the direct relationship between source width and signals reaching the ears is presented. For this task, the radiation characteristics of 10 acoustical instruments are recorded. By means of a simplification model, ear signals for 384 listening positions are calculated, neglecting room acoustical influences. Then, physical measures derived from the field of psychoacoustics and subjective room acoustics, are adapted to an anechoic environment. From these measures the actual source extent is predicted. Assuming that the perceived and the actual physical source extent largely coincide, these predictors give clues about the ear signals necessary to create the impression of a certain source width. This knowledge can be utilized for control over apparent source width in audio systems by considering the ear signals, instead of channel signals. It is an attempt at answering the question how perceived source extent is related to physical sound field quantities. A preliminary state of this study has been presented in Ziemer [50].

2 Perception of Source Width

Spatial hearing has been investigated extensively by researchers both in the field of psychoacoustics and in subjective room acoustics. Researchers in the first area tend to make listening tests under controlled laboratory conditions with artificial stimuli, such as clicks, noise and Gaussian tones. They investigate localization and the *perception of source width*. Researchers from the field of subjective room acoustics try to find correlations between sound field quantities in room impulse responses and sound quality judgments reported by expert listeners. Alternatively, they present artificial room acoustics to listeners, i.e. they use loudspeaker arrays in anechoic chambers. They observed that reflections can create the impression of a source that sounds even wider than the physical source extent. This auditory impression is referred to as *apparent source width*. Results from both research fields are addressed successively in this section.

2.1 Perceived Source Width in Psychoacoustics

Spatial hearing has been investigated mostly with a focus on sound source localization. Blauert [6] is one of the most comprehensive books about that topic. The localization precision lies around 1° in the frontal region, with a localization blur of about ±3.6°. Localization cues are contained in the head-related transfer function

(HRTF). It describes how a sound signal changes from the source to the ears. Monaural cues like overall volume and the distribution of spectral energy mainly serve for distance hearing. The further the source, the lower the volume. Due to stronger attenuation of high frequencies in the air, distant sources sound more dull than proximate sources. Furthermore, low frequencies from behind easily diffract around the pinnae. For high frequencies, the pinnae create a wave shadow. So the spectral energy distribution also helps for localization in the median plane. Binaural cues are interaural time differences (ITD) and interaural level differences (ILD) of spectral components. In dichotic playback, interaural phase differences (IPD) can be created without introducing ITD. Using forced-choice listening tasks and magnetoencephalography, Ross et al. [42] could prove, both behavioristically and neurally, that the human auditory system is sensitive to IPD below about 1.2 kHz.

Blauerte considers the localization blur the just noticeable difference (JND) in location whereas Zwicker and Fast1 consider it as precision with which the location of one stationary sound source can be given.[1] Both interpretations allow to hypothesize that the localization blur is related to width perception. The inability to name one specific angle as source angle may be due to the perception of a source that is extended over several degrees.

It is clear, however, that source localization and the perception of source width are not exactly the same. Evidence for this is the precedence effect which is sometimes referred to as *Haas effect* or *law of the first wavefront*.[2] The first arriving wave front is crucial for localization. Later arriving reflections hardly affect localization but can have a strong influence on the perceived source extent. Only a few authors investigated perceived source extent of the direct sound in absence of reflections. Hirvonen and Pulkki [24] have investigated the perceived center and spatial extent under anechoic conditions with a 45°-wide loudspeaker array consisting of 9 speakers. Through these, one to three non-overlapping, consecutive narrow-band noises were played by each speaker. The signals arrive simultaneously at a sweet-spot to minimize ITD and bias that results from the precedence effect. All loudspeakers were active in all runs. In all cases the perceived width was less than half the actual extent of the loudspeaker array. The authors were not able to predict the perceived width from the distribution of signals over the loudspeaker array. Investigating the relationship between perceived source width and ear signals, instead of loudspeaker signals, might have disclosed quantitative relationships. Furthermore, it might be difficult for a subject to judge the width of a distributed series of noise because such a signal is unnatural and not associated to a known source or a previously experienced listening situation. Natural sounds may have led to more reliable and predictable results. However, based on their analysis of

[1]Cf. Blauert [6], pp. 37f and Zwicker and Fastl [56], p. 309.

[2]See e.g. Haas [21], Blauert and Cobben [8].

channel signals they can make the qualitative statement that the utilized frequency range seems to have a strong impact on width perception.[3]

Potard and Burnett [39] found that "shapes", i.e. constellations of active loudspeakers, could be discriminated in the frontal region in cases of decorrelated white noise and 3 kHz high-pass noise in 42.5 and 41.4 % of all cases. Neither were subjects able to perform this task with 1 kHz low-pass noise and blues guitar, nor were they able to discriminate shapes in the rear for any kind of tested signal. The authors point out that perception of width and identification of source shape are highly dependent on the nature of the source signal. Furthermore, they observed that 70.4 % of all subjects rated a set of decorrelated sources more natural than a single loudspeaker for naturally large auditory events like crowd, beach etc. The findings that shapes of high-pass noise were discriminated better than shapes of low-pass noise underlines the importance of high-frequency content for the recognition of shapes. It could mean that ILD play a crucial role for the recognition of shapes. ILD mainly occur at high frequencies whereas low-pass noise mainly created IPD. The fact that high pass noise was discriminated better than blues guitar could furthermore denote that continuous sounds contain more evaluable information than impulsive sounds. The observation that only shapes in the frontal region could be discriminated may imply that experience with visual feedback improves the ability to identify constellations of sound sources. However, these assumptions are highly speculative and need to be confirmed by further investigations.

These two experiments demonstrate that subjects fail to recognize source width or shapes of unnaturally radiating sources, i.e. loudspeakers. Furthermore, mostly unnatural sounds are used, i.e. sounds that are not associated to a physical body, in contrast to the sound of musical instruments. In these two investigations loudspeaker signals are controlled. Control over the sound that actually reaches the listeners' ears might reveal direct cues concerning the relationship between the sound field and the perceived source width. Like blauerte states: "The sound signals in the ear canals (ear input signals) are the most important input signals to the subject for spatial hearing."[4] The investigation presented in Sect. 4 follows this paradigm, not controlling source signals but investigating what actually reaches the listeners' ears. The source signals are notes, played on real musical instruments including their natural sound radiation characteristics. Such signals are well-known to human listeners and associated with the physical extent of the instrument.

In many situations in which the listener is far away from the source, the physical source width is less than the localization blur. This is the case for most seats in concert halls for symphony music and opera. Here, the room acoustics, i.e. reflections, play a larger role for the auditory perception of source extent than the

[3]The complete investigation is documented in Hirvonen and Pulkki [24]. Contrary to width, they succeeded to replicate perceived source center by different adaptations of Raatgever's frequency weighting function.

[4]Blauert [6], p. 51.

direct sound. On the other hand, the radiation characteristics of sound sources have an immense influence on the room response. Apparent source width in room acoustics is discussed in the following.

2.2 Apparent Source Width in Room Acoustics

In the context of concert hall acoustics many investigations have been carried out to find relationships between physical sound field parameters and (inter-)subjective judgments about perceived source extent or overall sound quality. Since our acoustic memory is very short,[5] a direct comparison between listening experiences in different concert halls is hardly possible. Hence, listening tests have been conducted with experts, like conductors and music critics, who have long-term experience with different concert halls. Another method is to present artificially created and systematically altered sound fields or even auralize the complete room acoustics of concert halls. An overview about subjective room acoustics can be found in Beranek [4] and Gade [18].

In the context of subjective room acoustics, the apparent source width (ASW) is often defined as the auditory broadening of the sound source beyond its optical size.[6] Most authors agree that ASW is especially affected by direct sound and early reflections, arriving within the first 50–80 ms. Other terms that are used to describe this perception are *image* or *source broadening*, *subjective diffuseness* or *sound image spaciousness*.[7] All these terms are treated as the same in this chapter. The term *perceived source extent* is used to describe the auditory perception regardless of the quantities or circumstances that cause this impression.

The early lateral energy fraction ($\mathrm{LEF_{E4}}$) is proposed as ASW measure in international standards. It describes the ratio of lateral energy to the total energy at a receiver position $\mathbf{r}$ like[8]

$$\mathrm{LEF_{E4}}(\mathbf{r}) = \frac{\int_{t=5\,\mathrm{ms}}^{80\,\mathrm{ms}} p_8^2(\mathbf{r},t)\mathrm{d}t}{\int_{t=0}^{80\,\mathrm{ms}} p^2(\mathbf{r},t)\mathrm{d}t} \,. \tag{1}$$

Here, $p^2(\mathbf{r},t)$ is the squared room impulse response, measured by an omnidirectional microphone. The function $p_8^2(\mathbf{r},t)$ is the squared recording by a figure-of-eight-microphone whose neutral axis points towards the source. The subscript E stands for "early" and includes the first 80 ms. The subscript 4 denotes that the four octave bands around 125, 250, 500 and 1000 Hz are considered.

[5] See e.g. Gade [18], p. 304.

[6] See e.g. Blau [5], p. 720.

[7] See e.g. Yanagawa and Tohyama [47] and Yanagawa et al. [48].

[8] See e.g. Deutsches Institut f˙ur Normung [15], pp. 20f and Beranek [4], pp. 519 and 161.

The figure-of-eight microphone mainly records lateral sound whereas signals from the median plane largely cancel out. Hence, $\mathrm{LEF_{E4}}$ is the ratio of lateral to median sound or signal difference to signal coherence. The larger the value, the wider the expected ASW. In a completely diffuse field a value of $\mathrm{LEF_{E4}} = 0.33$ would occur.[9]

Beranek [4] found a significant negative correlation between ASW and the early interaural crosscorrelation ($\mathrm{IACC_{E3}}$). The subscript 3 denotes that the mean value of three octave bands around 500, 1000 and 2000 Hz is considered. $1 - \mathrm{IACC_{E3}}$ is also known as *binaural quality index* (BQI). BQI shows positive correlation to ASW. It is calculated from the $\mathrm{IACC_E}$, which is the maximum absolute value of the interaural crosscorrelation function (IACF) as measured from band passed portions of impulse response recordings with a dummy head:

$$\mathrm{IACF_E}(\mathbf{r}, \tau) = \frac{\int_{t=0}^{80\,\mathrm{ms}} p_\mathrm{L}(\mathbf{r}, t) p_\mathrm{R}(\mathbf{r}, t+\tau)\mathrm{d}t}{\sqrt{\int_{t=0}^{80\,\mathrm{ms}} p_\mathrm{L}^2(\mathbf{r}, t)dt \int_{t=0}^{80\,\mathrm{ms}} p_\mathrm{R}^2(\mathbf{r}, t)\mathrm{d}t}} \tag{2}$$

$$\mathrm{IACC_E}(\mathbf{r}) = \max|\mathrm{IACF_E}(\mathbf{r}, \tau)| \tag{3}$$

$$\mathrm{BQI}(\mathbf{r}) = 1 - \mathrm{IACC_{E3}}(\mathbf{r}) \tag{4}$$

The subscripts L and R denote the left and the right ear. The variable τ describes the time lag, i.e. the interval in which the interaural cross correlation is searched; $\tau \in (-1, 1)$ ms roughly corresponds to the ITD of a completely lateral sound. The IACC is calculated individually for each of the three octave bands. Their mean value is $\mathrm{IACC_{E3}}$. Beranek [4] found a reasonable correlation between LEF and BQI, which is not confirmed by other authors.[10] Ando even found neural correlates to BQI in the brainstem of the right hemisphere which is a strong evidence that the correlation of ear signals is actually coded and processed further by the auditory system.[11] It is conspicuous that two predictors of ASW—namely $\mathrm{LEF_{E4}}$ and BQI—consider different frequency regions. In electronically reproduced sound fields Okano et al. [37] have found that a higher correlation could be achieved when combining BQI with $G_\mathrm{E,low}$, the average early strength of the 125- and 250 Hz-octave band which is defined as

$$G_\mathrm{E,low}(\mathbf{r}) = 10\,\mathrm{lg}\frac{\int_{t=0}^{80\,\mathrm{ms}} p^2(\mathbf{r}, t)\mathrm{d}t}{\int_{t=0}^{\mathrm{dir}} p_\mathrm{ref}^2(t)\mathrm{d}t}\ . \tag{5}$$

$G_\mathrm{E,low}$ is the ratio between sound intensity of a reverberant sound and the pure direct sound p_ref. lg is the logarithm to the base 10 and the denominator represents

[9] According to Gade [18], p. 309.

[10] Cf. Beranek [4], p. 528 versus Blau [5] and Gade [18], p. 310.

[11] See Ando [2], p. 5.

the integrated squared sound pressure of the pure direct sound, which is proportional to the contained energy. The finding that strong bass gives rise to a large ASW even when creating coherent ear signals is not surprising. In nature only rather large sources tend to radiate low-frequency sounds to the far field. Here, the wavelengths are so large that barely any interaural phase- or amplitude differences occur. From psychoacoustic investigations it is known that monaural cues help for distance hearing. And distance, of course, strongly affects source width if we consider the relative width in degrees from a listener's point of view.

An alternative measure that includes the enlarging effect of strong bass frequencies is the *interaural difference*

$$\mathrm{IAD}(\mathbf{r}) = 10\,\mathrm{lg}\left(\frac{\mathrm{eq}(p_\mathrm{L}(\mathbf{r},t) - p_\mathrm{R}(\mathbf{r},t))^2}{p_\mathrm{L}^2(\mathbf{r},t) + p_\mathrm{R}^2(\mathbf{r},t)}\right). \tag{6}$$

This measure is proposed in Griesinger [20]. Basically, it is the difference signal of the squared dummy head recordings divided by the sum of their squared signals. The signal difference between the two dummy head ears is similar to a recording with a figure-of-eight microphone, and quantifies lateral sound energy. Their sum approximate an omnidirectional recording. Here, phase inversions cancel out and the mono component of the sound field is quantified. The factor eq stands for an equalization of the difference signal. Frequencies below 300 Hz are emphasized by 3 dB per octave. Due to their large wavelengths, bass frequencies hardly create interaural phase differences, even in a reverberant sound field. Consequently, a strong bass reduces values for $\mathrm{LEF}_{\mathrm{E4}}$, which contradicts the listening experience. This is probably the reason why the BQI does not consider such low frequencies. The equalization in the IAD counteracts this false trend. Unfortunately, the paper does not report any experience with this measure and its relationship to ASW.

Another approach to take the widening effect of low frequencies into account is to consider the width of the major IACF peak (W_{IACC}). Low frequencies tend to create wide IACF peaks, because small time lags barely affect phase. So W_{IACC} is related to the distribution of spectral energy. Shimokura et al. [44] even states that W_{IACC} is correlated to the spectral centroid of a signal. In Ando [2], it is described that a combination like

$$\mathrm{ASW}_{\mathrm{pre}} = \alpha(\mathrm{IACC})^{3/2} + \beta(W_{\mathrm{IACC}})^{1/2} \tag{7}$$

yields a very good prediction of ASW of band pass noise, if α and β are calculated for individuals.[12] For multi-band noise, the binaural listening level (LL) is an important additional factor.

Of all objective parameters that are commonly measured in room acoustical investigations, the $\mathrm{IACC}_{\mathrm{E}}$, and the strength G belong to the quantities that are most sensitive to variations of the sound radiation characteristics. In Martin et al. [35],

[12]See Ando [2], p. 130ff.

acoustical parameters are measured for a one source-receiver constellation but with two different dodecahedron loudspeakers. Although both loudspeakers approximate an omnidirectional source, deviations of G and BQI are larger than the just noticeable difference, i.e. they are assumed to be audible. In their experiment, this is not the case for $\mathrm{LEF_{E4}}$. This is probably the case because $\mathrm{LEF_{E4}}$ mainly considers low frequencies. Dodecahedron loudspeakers approximate an omnidirectional source much better at low frequencies than at high frequencies. Although good correlations between reported ASW and measured BQI could be found in many studies, this measure is not always a reliable predictor. It has been found that BQI tends to have massive fluctuation even when only slightly moving the dummy head. The same is true for $\mathrm{LEF_{E4}}$. These fluctuations are not in accordance with listening experiences.[13] When sitting in one concert hall seat and slightly moving the head, the ASW does not change as much as the BQI and the $\mathrm{LEF_{E4}}$ indicate. From a perceptual point of view, an averaging of octave bands is questionable, anyway. The auditory system rather averages over critical bands which can be approximated better by third-octave bands. Consequently, these measures are not valid for one discrete listening position $\mathbf{r}$. Their spatial averages over many seats rather give a good value for the overall width impression in the concert hall under consideration. This finding has been confirmed partly in Blau [5]. In listening tests with synthetic sound fields, the author could not find an exploitable correlation between ASW and BQI when considering all investigated combinations of direct sound and reflection. Only after eliminating individual combinations a correlation could be observed. He could prove that the fluctuations of BQI over small spatial intervals is not the only reason for the low correlation. He observed a higher correlation between ASW and $\mathrm{LEF_{E4}}$, which could explain $R^2 = 64$ % of the variance with one pair of reflections and $R^2 = 88$ % with multiple reflections. Assuming that frequencies above 1 kHz as well as the delay of single reflections may play a considerable role, Blau [5] proposed

$$\mathrm{RL_E} = 10\,\mathrm{lg}\frac{\sum_{i=1}^{n} a_i \sin\alpha_i E_i}{E_D + \sum_{i=1}^{n}(1 - a_i \sin\alpha_i)E_i} \tag{8}$$

as measure for ASW.[14] Here, i is the time window index. Time windows have a length of 2 ms and an overlap of at least 50 %. The upper bound n is the time window that ends at 80 ms. The weighting factor $a_i = 1 - e^{-t_i/15\,\mathrm{ms}}$ is an exponentially growing factor to emphasize reflections with a larger delay. α_i is the dominant sound incidence angle in the ith time window. It is estimated from an IACF of the low-passed signals weighted by a measure of ILD. E_D is the energy of the direct sound, E_i is the reflected energy contained in the ith time window.

[13]For details on the spatial fluctuations of BQI and LEFE4 refer to de Vries et al. [14].

[14]See Blau [5], p. 721.

The RL_E explained 89–91 % of the variance. It could be proved that the BQI changes when exciting the room using continuous signals instead of an impulse.[15] This finding may indicate that this measure cannot be applied to arbitrary signals. On the other hand, Potard and Burnett [39] already found out that the discrimination of shapes works with continuous high-pass noise but not with blues guitar. Likewise, width perception could be different for impulsive and continuous signals, so a measure for ASW does not necessarily need to have the same value for an impulse and a continuous signal. In the end, the BQI does not claim to predict ASW under conditions other than concert hall acoustics. It considers an omnidirectional impulse and does neither make a clear separation between direct sound and reflections nor does it take the radiation characteristics of sources into account. The radiation characteristics have a strong influence on the direct sound and the room acoustical response.

In Shimokura et al. [44], the IACC of a binaural room impulse response is differentiated from an $IACC_{SR}$ of an arbitrary source signal. They propose some methods to translate $IACC_{SR}$ to IACC, which are out of scope of this chapter. The authors convolve dry signals of musical instruments with binaural room impulse responses to investigate the relationship between perceived width and $IACC_{SR}$ with different signals. This way, different performances in the same hall can be compared as well as the same performance in different halls. By multiple linear regression the authors tried to predict reported diffuseness (SV) from descriptors of the signals' autocorrelation functions (ACFs) by

$$\mathrm{SV}(\mathbf{r}) = a\mathrm{IACC}(\mathbf{r}) + b\tau_e + cW_{\phi(0)}(\mathbf{r}) + d\ . \tag{9}$$

Here, $W_{\phi(0)}$ is the width of the first IACF peak and τ_e is the duration until the envelope of the ACF falls by 10 dB. It is 0 for white noise and increases when decreasing the bandwidth and converges towards ∞ for a pure tone. The contribution of IACC was significant for eight of nine subjects, whereas the contribution of τ_e and $W_{\phi(0)}$ was only significant for four and two of nine. Consequently, the multiple linear regression failed to explain SV of all subjects. Just as in the approach of Ando [2], Eq. 7, the factors a, b and c had to be adjusted for each individual. Shimokura et al. [44] observed that W_{IACC} was only significant for one individual subject which contradicts the findings of Ando [2]. Both approaches explain subjective ratings on the basis of objective parameters but their findings do not exhibit intersubjective validity.

Based on psychophysical and electrophysiological considerations, Blauert and Cobben [8] proposed a running cross correlation (RCC) of recorded audio signals

[15] See Mason et al. [36].

$$\mathrm{RCC}(\mathbf{r}, t, \tau) = \int_{-\infty}^{t} q_\mathrm{L}(\mathbf{r}, \delta) q_\mathrm{R}(\mathbf{r}, \delta + \tau) G(\mathbf{r}, t - \delta) \mathrm{d}\delta \, . \tag{10}$$

Here, q is the recorded signal p after applying a half-wave rectification and a smoothing in terms of low-pass filtering. The RCC is a function of time and lag, so it yields one cross correlation function for each time step. $G(\mathbf{r}, t - \delta)$ is a weighting function to attenuate past values

$$G(s) = \begin{cases} e^{\frac{-s}{5\,\mathrm{ms}}} & \text{for } \begin{cases} s \geq 0 \\ s < 0 \end{cases} \\ 0 \end{cases} . \tag{11}$$

The RCC produces peaks that are in fair agreement with lateralization judgments and the precedence effect, i.e. a dominance of the first wavefront. But the authors emphasize the need for improvements.

Yanagawa and Tohyama [47] conducted an experiment with a leading sound and a delayed copy of it, simulating direct sound and one reflection. They found that the interaural correlation coefficient (ICC) is a better estimator of source broadening than BQI. The ICC equals the ICCF, Eq. 2, when τ is chosen to be 0. Lindemann [33] uses the same measure but divides the signal into several frequency bands. He hypothesizes that small differences between the perceived location of frequency bands are the reason for subjective diffuseness.

Blauert and Lindemann [9] found evidence that early reflections with components above 3 kHz create an image expansion. But Bradley et al. [10] have found that late arriving reflections may again diminish ASW. However, the idea of ASW is that a listener is rather far away from the source. Consequently, the original width of a musical instrument is in the order of one degree or less. This original sound source is "extended" due to a decorrelation of ear signals which are caused by unsymmetrical reflections. But when being close enough to a musical instrument, it does have a notable width of many degrees. This width can be heard. In proximity to a source, direct sound already creates decorrelated signals at both ears. This decorrelation mainly results from the frequency- and direction-dependent radiation characteristics of musical instruments. Decorrelation of stereo and surround channels is common practice in music production to achieve the sensation of a broad sound source. In ambisonics and wave field synthesis, complex source radiation patterns are synthesized to create this impression. Source width in music production is discussed in the following section.

3 Source Width in Music Production

Perceived source width is of special interest in music production. In text books for recording, mixing and mastering engineers, spaciousness plays a major role. In the rather practical book written by Levinit [32], a chapter about recording tips and

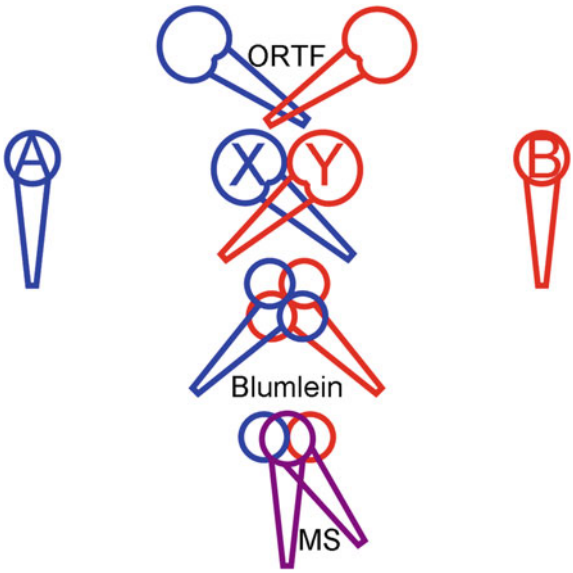

Fig. 1 Common stereo recording techniques

tricks has a section named "Making Instruments Sound Huge". Likewise, the audio engineer Kaiser [28] points out that the main focus in mastering lies in the stereo width, together with other aspects, such as loudness, dynamics, spaciousness and sound color.[16]

Probably by hearing experience, rather than due to fundamental knowledge of psychoacoustics and subjective room acoustics, sound engineers have found several ways to capture the width of musical instruments via recording techniques or to make them sound larger by pseudo-stereo methods. These are discussed in this section, followed by methods of source broadening in ambisonics and wave field synthesis application.

3.1 Source Width in Stereo and Surround

For recorded music, several microphoning techniques have been established. In the far field, they are used to capture the position of instruments in an ensemble and to record different portions of reverberation. In the near field, they capture the width of a solo instrument to a certain degree. Figure 1 shows some common stereo microphone techniques, namely A-B, Blumlein, mid-side stereo (MS), ORTF and X-Y. They are all based on a pair of microphones. The directivity of the microphones is depicted here by the shape of the head: omnidirectional, figure-of-eight

[16]See Kaiser [28], e.g. p. 23 and p. 40.

and cardioid. The color codes to what stereo channel the signal is rooted. Blue means left channel, red means right channel and violet denotes that the signal is routed to botch channels. Directional microphones that are placed closely together but point at different angles create mainly inter-channel level differences (ICLDs). This is the principle of X-Y recording. In A-B-recording, a large distance between microphones creates additional inter-channel time differences (ICTDs). So the recording techniques create systematically decorrelated stereo signals. The Blumlein recording technique creates even stronger ICLDs for frontal sources but more ambient sound or rear sources are recorded as well. In MS, sound from the neutral axis of the figure-of-eight microphone is only recorded by the omnidirectional microphones. It is routed to both stereo channels. The recording from the figure-of-eight microphone mainly captures lateral sound incidence and is added to the left and subtracted from the right channel. MS recording is quite flexible because the amplitude ratio between the monaural omnidirectional (mid-component) and the binaural figure-of-eight recording (side-component) can be freely adjusted. In all recording techniques, the degree of ICLD and ICTD depends on the position and radiation patterns of the source as well as on the amount and characteristics of the recording room reflections. More details on the recording techniques are given e.g. in Kaiser [27] and Friedrich [17].[17] It is also common to pick up the sound of musical instruments at different positions in the near field, for example with one microphone near the neck and one near the sound hole of a guitar. This is supposed to make the listener feel like being confronted with an instrument that is as large as the loudspeaker basis or like having the head inside the guitar.[18] When a recording sounds very narrow, it can be played by a loudspeaker in a reverberation chamber and recorded with stereo microphone techniques.[19] This can make the sound broader and more enveloping.

Recording the same instruments twice typically yields a stronger and, more importantly, dynamic decorrelation. Slight differences in tuning, timing, articulation and playing technique between the recordings occur. As a consequence, the relation of amplitudes and phases, transients and spectra changes continuously. These recordings are hard-panned to different channels, typically with a delay between them.[20] This overdubbing technique occurred in the 1960s.[21] Virtual overdubbing can be performed if the recording engineer has only one recording.[22] Adding one chorus effect to the left and a phase-inverted chorus to the right channel creates a dynamic decorrelation. In analog studios, artificial double tracking (ADT) was applied to create time-variant timing-, phase- and frequency differences between

[17]See especially Kaiser [27], pp. 33–43 and Friedrich [17], Chap. 13.

[18]This promise is made in Levinit [32], p. 157.

[19]See e.g. Faller [16].

[20]This is especially done for guitar and some vocal parts, see e.g. Kaiser [26], p. 116f and p.127 and Hamidovic [22], p. 57.

[21]See e.g. Maempel [34], p. 236.

[22]See e.g. Cabrera [11].

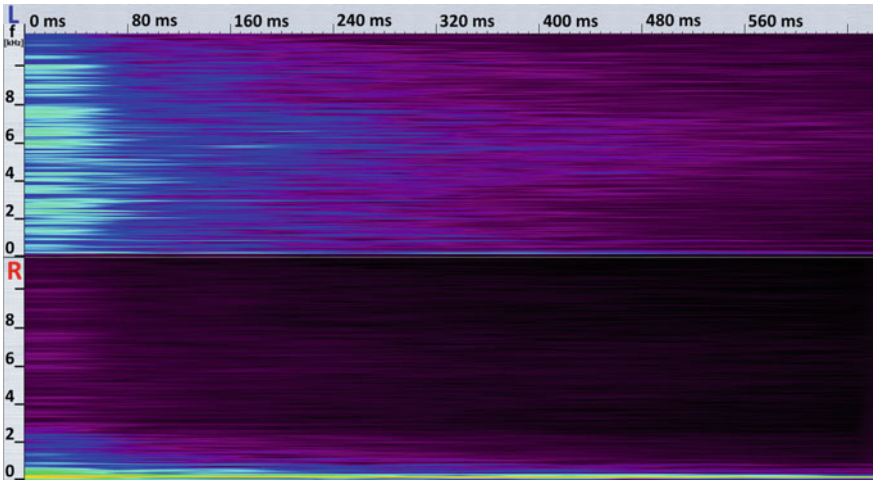

Fig. 2 Pseudostereo by high-passing the *left* and low-passing the *right* channel

channels. Here, a recording is re-recorded, using wow and flutter effects to alter the recording tape speed dynamically.

For electric and electronic instruments as well as for recorded music, several pseudostereo techniques are commonly applied to create the impression of a larger source. An overview of pseudo-stereophony techniques is given in Faller [16]. For example, sound engineers route a low-passed signal to the left and a high-passed signal to the right loudspeaker to increase the perceived source width as illustrated in Fig. 2. All-pass filters can be used to create inter-channel phase differences (ICPD) while maintaining a flat frequency response. Some authors report strong coloration effects, others less.[23] Usually, filters with a flat frequency and a random phase response are chosen by trial-and-error. Another method is to apply complementary comb filters[24] as indicated in Fig. 3. These create frequency-dependent ICLDs. Played back in a stereo setup, these ICLDs create ILDs but mostly to a lower degree, because both loudspeaker signals reach both ears. The ILDs are interpreted as different source angles by the listener. But, as long as the signals of the spatially spread frequency bands share enough properties, they remain fused. They are not heard as different source angles but as one spread source. Schroeder [43] investigated which sound parameters affect spatial sound impressions in headphone reproduction. He comes to the conclusion that ILD of spectral components have a greater effect on the perception of source width than IPD. Often, an ICTD between 50 and 150 ms is used to create a wide source. Sometimes, the delayed and attenuated copy of the direct sound is directly routed to the left channel and phase-inverted for the right. Applying individual filters or compressors for each

[23]See e.g. Cabrera [11] and Zotter and Frank [54] versus Faller [16].

[24]See e.g. Cabrera [11] and Kaiser [28], p. 154.

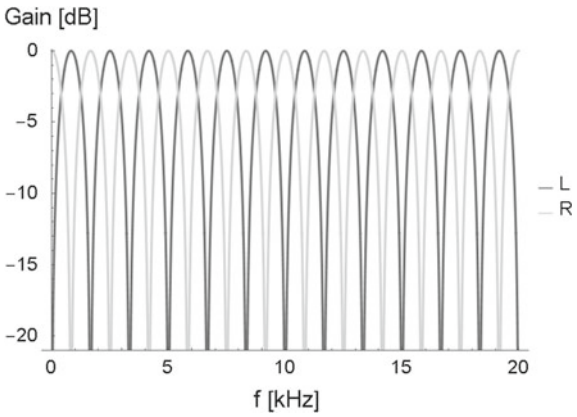

Fig. 3 Pseudostereo by applying complementary comb filters on the left and the right channel

channel is common practice, as well as creating a MS stereo signal and compressing or delaying only the side-component.[25] Likewise, it is very common to apply complementary equalizers to increase separation between instruments in the stereo panorama or to pan the reverb to a location other than the direct sound.[26] One additional way to create a higher spaciousness is to use a Dolby surround decoder on a stereo signal. This way, one additional *center channel* and one *rear channel* are created. These can be routed to different channels in a surround setup. The first is basically the sum of the left and the right channel whereas the latter is their difference, which is high-passed and delayed by 20–150 ms. This effect is called *magic surround*.[27] A general tip for a natural stereo width is to make bass frequencies most mono, mid-range frequencies more stereo and high frequencies most stereo,[28] i.e. with an increasing decorrelation of channels.

All of the named pseudo-stereo techniques are based on the decorrelation of loudspeaker signals. The idea is that the resulting interaural correlation is proportional to channel correlation. There are only few monaural methods to increase perceived source width. One practice is to simply use a compressor. The idea is inspired by the auditory system which, because of the level-dependent cochlear gain reduction, in fact operates as a 'biological compressor'. So a technical signal compressor creates the illusion that a source is very loud, and consequently very proximate to the listener. Naturally, proximate sources are wider, i.e. they are spread over more degrees from the listeners' point of view. Especially low frequencies should be compressed with a high attack time.[29]

[25]See Hamidovic [22], p. 57 and Kaiser [28], p. 152 and 156.

[26]See Kaiser [26], p. 50 and pp. 57f.

[27]See e.g. Faller [16] and Slavik and Weinzierl [45], p. 624.

[28]See Kaiser [28], pp. 148f.

[29]See e.g. Levinit [32], p. 158 and Rogers [41], p. 35.

Faller [16] proposes two additional pseudo-stereophony methods. The first is to compare a mono recording to a modern stereo mix and then create the same ICTD, ICLD and ICC for every subband. The second is to manually select auditory events in the spectrogram of the mono file and apply panning laws to spread instruments over the whole loudspeaker basis. Zotter and Frank [54] systematically alter inter-channel amplitude or phase differences of frequency components to increase stereo width. They found that the inter-channel cross correlation (ICCC) is approximately proportional to IACC in a range from $\mathrm{IACC}_u = 0.3$ to $\mathrm{IACC}_o = 0.8$. For both amplitude and phase alterations, they observe audible coloration.[30] Laitinen et al. [31] utilize the fact that in reverberant rooms, in contrast to anechoic conditions, the interaural coherence decreases with increasing distance to a sound source. This is not surprising as the direct-to-reverberant energy ratio (D/R ratio) decreases. The direct sound, which creates relatively high interaural coherence, is attenuated whereas the intensity of the relatively diffuse reverberance remains the same. Likewise, loudness and interaural phase coherence decreases with increasing distance to the source. They present formulas to control these three parameters. Gain factors are derived simply from listening to recreate the impression of three discrete distances. Control over perceived source distance might be related to perceived source extent.

In recording studios, a typical analyzing tool is the so-called *phase scope*, *vectorscope* or *goniometer*, plotting the values of the last x samples of the left versus the right channel as discontinuous Lissajous figures and additionally giving the inter-channel cross correlation coefficient.[31] This analysis tool is applied to monitor stereo width. It is illustrated in Fig. 4. The inter-channel cross correlation coefficient informs about mono compatibility. A negative correlation creates destructive interference when summing the stereo channel signals to one mono channel. When the left and right channel play the same signal, the goniometer shows a straight line. If amplitude differences occur, the line is deflected towards the channel with the louder signal. The more complicated the relation between the channel signals, the more chaotic the goniometer plot looks.

For surround systems with 5 or more channels, multi directional amplitude panning (MDAP) has been proposed. The primary goal of MDAP is to solve the problem of discontinuity: When applying amplitude based panning between pairs of loudspeakers, the perceived width of phantom sources is larger in the center and becomes more narrow for phantom source positions that are close to one of the loudspeakers. To increase the spread of lateral sources at least one additional speaker is activated. The principle is illustrated in Fig. 5. A target source width is chosen. It has to be at least the distance of two neighboring loudspeakers. One phantom source is panned to the left end of the chosen source extent, one phantom

[30]See Zotter and Frank [54] for details on their channel decorrelation methods and their investigations of IACC and sound coloration.

[31]See e.g. Kaiser [28], pp. 48ff although the meaning of the correlation coefficient is obviously misunderstood by this practician.

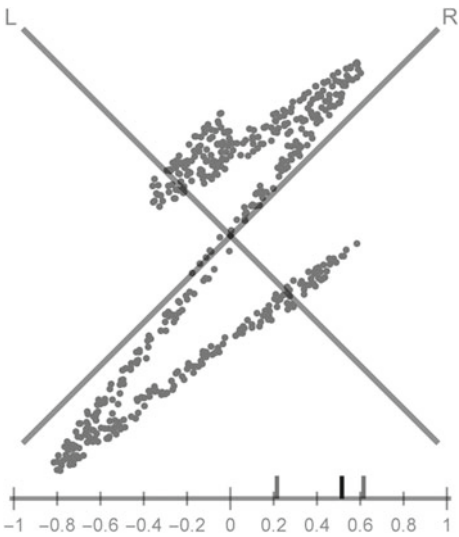

Fig. 4 Phase space diagram (*top*) and correlation coefficient (*bottom*) as objective measures of stereo width and mono compatibility

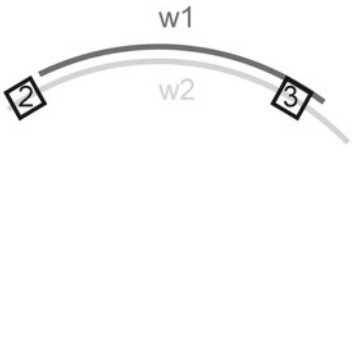

Fig. 5 Multi dimensional amplitude panning for different source widths

source is panned to the right end. For the illustrated source $w1$, loudspeakers 2, 3 and 4 are active. Source $w2$ has the same central source angle but a wider source extent. Here, loudspeaker 1 is additionally active.

3.2 Source Width in Ambisonics

Ambisonics started as microphone and playback technique in the 1970s. Pioneering work has been done by Gerzon.[32] The basic two-dimensional ambisonics recording technique is illustrated in Fig. 6. It is referred to as *first order ambisonics*.

[32]See e.g. Gerzon [19].

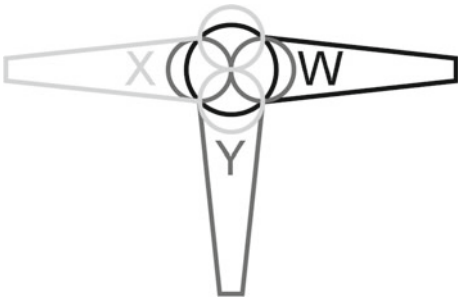

Fig. 6 First order ambisonics recording technique

One pressure microphone W and two perpendicular pressure gradient microphones X and Y are used. In the three-dimensional case, an additional figure-of-eight microphone captures the pressure gradient along the remaining axis, referred to as *B-Format* or *W, X, Y, Z*. Three-dimensional audio is out of scope of this chapter.

In contrast to conventional stereo recording techniques, the signals are not directly routed to discrete loudspeakers. They rather encode spatial information, namely the pressure distribution on a circle. The three microphones perform a truncated circular harmonic decomposition of the sound field at the microphone position. The monopole recording W gives the sound pressure at the central listening position p_0, i.e. the circular harmonic of 0th order. It is routed directly to the zeroth channel, i.e.

$$\text{ch0} = \frac{W}{\sqrt{2}} \,. \tag{12}$$

Recordings X and Y are the pressure gradients along the two spatial axes, i.e. 1st order circular harmonics. They can be approximated by

$$\text{ch1} = X \approx p_c(0) - p_c(\pi) \tag{13}$$

and

$$\text{ch2} = Y \approx p_c\left(\frac{\pi}{2}\right) - p_c\left(\frac{3\pi}{2}\right) \,. \tag{14}$$

Here, $p_c(\phi)$ are omnidirectional recordings of microphones that are distributed along a circle with a small diameter. Higher order encoding can be performed with more pressure receivers. For an encoding of order n, $4n + 1$ pressure receivers are necessary. Figure 7 illustrates ambisonics recordings of different orders for the same wave field. Recordings from microphones at different angles are combined like

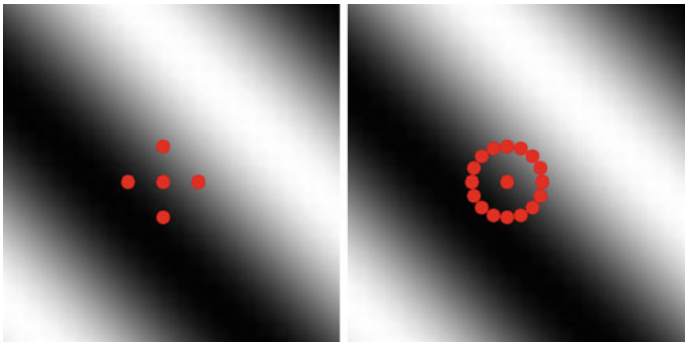

Fig. 7 1st order (*left*) and 4th order (*right*) ambisonics recording of a plane wave

$$\text{ch3} \approx p_c(0) - p_c\left(\frac{\pi}{2}\right) + p_c(\pi) - p_c\left(\frac{3\pi}{2}\right) \tag{15}$$

and

$$\text{ch4} \approx p_c\left(\frac{\pi}{4}\right) - p_c\left(\frac{3\pi}{4}\right) + p_c\left(\frac{5\pi}{4}\right) - p_c\left(\frac{7\pi}{4}\right) . \tag{16}$$

Figure 8 illustrates the circular harmonics. Their superposition yields the *n*th order approximation of the sound field along the circle. The first order approximation yields a cardioid. The maximum points at the incidence angle of the wave front. The lobe is rather wide. In contrast to that, the maximum of the 4th order approximation is a relatively narrow lobe that points at the incidence angle of the wave front. However, several sidelobes occur. The order gives the precision with which the sound field is encoded. For one plane wave, the first order approximation already yields the source angle. For superimposed sound fields with several sources and complicated radiation patterns, a higher order is necessary to encode the sound field adequately. However, a finite order might always contain artifacts due to sidelobes.

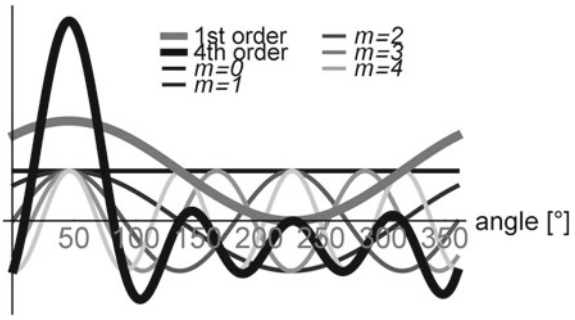

Fig. 8 Circular harmonics of order 0 and 1 are encoded in 1st order ambisonics. In 4th order ambisonics, additional circular harmonics of order 2, 3 and 4 are necessary

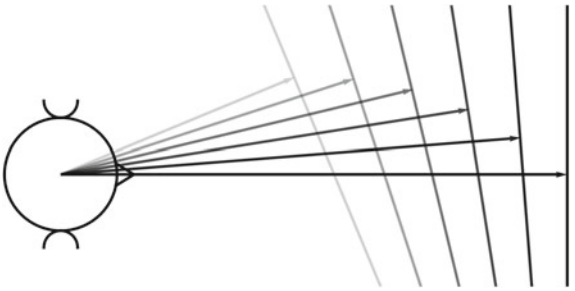

Fig. 9 Phantom source widening in ambisonics by synthesizing frequency dispersed source positions. Different frequency regions are indicated by different gray levels

Ambisonics decoders use different strategies to synthesize the encoded sound field at the central listening position. This is either achieved by the use of projection or by solving a linear equation system that describes the relationship between loudspeaker position, wave propagation and the encoded sound field on a small circle around the central listening position. Ambisonics decoders are out of scope of this chapter. An overview can be found e.g. in Heller [23].

Zotter et al. [55] propose a method which is related to the idea of a frequency-dependent MDAP. In an ambisonics system, frequency regions are not placed at the same source position but spread over discrete angles. In a way, this is a direct implementation of the hypothesis that has been formulated by Lindemann [33] who believes that deviant source localizations of different frequency bands is the reason for subjective diffuseness. The principle is illustrated in Fig. 9. In their listening test, the perceived source extent, reported by 12 subjects, correlated with the BQI when increasing the time lag to $\tau = 2$ ms.[33]

Another principle is tested in Potard and Burnett [40]. They synthesize 6 virtual point sources with 4th order ambisonics. The virtual source positions are spread over different angles. White noise is divided into three frequency bands. The signal for each virtual point source is composed of decorrelated versions of these frequency bands. The decorrelation is achieved by all pass filters. Then, they mix each frequency band of the original source signal with the decorrelated version. With the mixing ratio ξ and the distribution of the virtual point sources, they try to control the source width of each frequency region. The perceived source extents reported by 15 subjects are in fair agreement with the intended source extents. Unfortunately, no systematic alteration of virtual source spread and degrees of decorrelation are presented in their work.

The authors in Laitinen et al. [30] propose an implementation of directional audio coding (DirAC) in ambisonics. A premise of their approach is that the human auditory system perceives exactly one direction and one source extent for each frequency band in each time frame. From an ambisonics recording they derive the source angle and its diffuseness in terms of short-term fluctuations or uncertainty.

[33]Their approach and experiment are documented in Zotter et al. [55]. The information that thetime lag was increased cannot be found in the paper; it was given verbally at the conference.

The source angle is created by ambisonics decoding. Diffuseness is created by decorrelated versions that are reproduced by different loudspeakers. In a listening test with 10 subjects, they found that localization and sound quality were very good with their approach. For future research, they propose to investigate the perceived source extent in more detail.

Just as in stereo, the presented ambisonics approaches either aim at controlling the signals at discrete channels or at controlling the spatial spread of virtual sources. Focusing on the sound field at the listening position might reveal a deeper insight into the relationship between ear signals and the perception of width. This is not the case for all wave field synthesis techniques. These are discussed in the following.

3.3 Source Width in Wave Field Synthesis

Wave field synthesis is based on the idea that the sound field within an enclosed space can be controlled by signals on its surface. An overview of its theory and application can be found in Ziemer [51]. Typically, wave fronts of static or moving virtual monopole sources or plane waves are synthesized in an extended listening area. With this procedure, listeners experience a very precise source location which stays stable, even when moving through the listening area. However, due to the simple omnidirectional radiation pattern, virtual sources tend to sound small. This observation called several researchers into action, trying to make sources sound larger, if desired.

Baalman [3][34] arranged a number of virtual point sources to form a sphere, a tetrahedron and an icosahedron, each with a diameter of up to 3.4 m. With this distribution of virtual monopole sources, she played speech and music to subjects. The shapes were perceived as being further away and broader than a monopole source. The most perceivable difference was the change in tone color. In her approach the perceived source width did not depend on the width of the distributed point sources. There are several potential reasons why her method failed to gain control over perceived source widths. One reason might be that the distributed point sources radiated the same source signal. No filtering or decorrelation was performed. Except for low frequencies, coherent sound radiation from all parts of a source body is rather unusual and does not create the perception of a large source width. Wave field synthesis works with exactly this principle; delayed and attenuated versions of the same source signal are played by a closely spaced array of loudspeakers to recreate the wave front of a virtual monopole source or plane wave. Thus, the difference between one virtual monopole and a spherical distribution of coherent virtual monopoles can only lie in synthesis errors and in comb filter effects that depend on the distance of the point sources. Another reason might have been that the distance between listeners and source was in all cases more than 3 m. So

[34]See Baalman [3], Chap. 7.

Fig. 10 Combined (*left*) and plain (*right*) multipoles of low orders

when measuring source width in degrees, the shapes are again relatively narrow in most trials.

In Corteel [12], the synthesized sources are no monopoles but circular harmonics of order 1–4 and some combinations of those, i.e. multipoles. Some exemplary radiation patterns are illustrated in Fig. 10. The paper focuses on the optimization of filters to minimize physical synthesis errors. It does not include listening tests that inform about perceived source extent. However, as soon as a multipole of low order is placed further than a few meters away from a listener, it barely creates interaural sound differences. The reason is that multipoles of low order are very smooth. Assuming a distance of 0.15 m between the ears, the angle between the ears and a complexly radiating point source at 3 m distance is about 2.8°. Only slight amplitude and phase changes occur over this angle width for low order multipoles. This can easily be seen in Fig. 10. For steep, sudden changes to occur within a few degrees, a very high order is necessary.

In Jacques et al. [25], single musical instruments or ensembles are recorded with a circular microphone array consisting of 15 microphones. They synthesize the recordings by means of virtual high order cardioid sources, pointing away from the origin, i.e. the original source point. This way, the radiation pattern is reconstructed to a certain degree. In a listening test, subjects were able to hear the orientation of a trumpet with this method. When synthesizing only one high order cardioid, many subjects had troubles localizing the source. This was, however, not the case when several high order cardioids reconstruct an instrument radiation pattern.

In Ziemer and Bader [53], the radiation characteristic of a violin is recorded with a circular microphone array which contains one microphone every 2.8°. The radiation characteristic is synthesized in a wave field synthesis system. This is achieved by simplifying the violin as complex point source. The physical approach is the same as in the present study and will be explained in detail in Sect. 4.2. The main aim of this paper is to utilize psychoacoustic phenomena to allow for physical synthesis errors while ensuring precise source localization and a spatial sound impression. In a listening test with 24 subjects, the recreated violin pattern could be localized better than a stereo phantom source with plain amplitude panning. Still, it was perceived as sounding more spatial.

The approach to model virtual sources with more complex radiation characteristics to achieve control over ASW is very promising. But it is necessary to create the cues that affect ASW. These cues are to be created by the virtual source and by synthesized reflections. But more important than the sound field at the virtual source position is the sound field at the ears of the listener. In the study that is

described in the following section, relationships between source width and the sound field at listening positions are investigated.

4 Sound Radiation and Source Extent

In this investigation the actual extent of the vibrating part of certain musical instruments is related to quantities of the radiated sound. Here, the focus lies on direct sound. The idea behind this procedure is straightforward: There must be evaluable quantities in the radiated sound that indicate source width because the auditory system has no other cues than these. As mentioned earlier, investigations which aimed at explaining perceived source width of direct sound by controlling signals of loudspeakers—instead of the signals at listeners' ears—did not succeed. But if we find parameters in the radiated sound that correlate with actual physical width we may have found the cues which the auditory system consults to render a judgment about source width. By controlling these parameters, more targeted listening tests can be conducted. Furthermore, when the relationship between audio signal and width perception is disclosed, it can be implemented as a tool for stereo, ambisonics and wave field synthesis applications to control perceived source extent.

This investigation is structured as follows: First, the setup to measure the radiation patterns of musical instruments is introduced and the examined instruments are listed. Then, the complex point source model is briefly described. The model is applied to propagate the instrumental sound to several potential listening positions. For these listening positions, physical sound field quantities are calculated. Basically, the quantities are taken from the field of psychoacoustics and subjective room acoustics. But they are adopted to free field conditions and instrumental sounds. The adopted versions are discussed subsequently. Finally, relationships between sound field quantities and the physical source extent are shown. It is demonstrated how a combination of two parameters can be used to predict the source extent. Although physical sound field quantities are put into relation with physical source extent, the findings allow some statements about psychoacoustics. So the results are discussed against the background of auditory perception. Potential applications and future investigations are proposed in the prospects section.

4.1 Measurement Setup

In an anechoic chamber a circular microphone array was installed roughly in the height of the investigated musical instruments. It contains 128 synchronized electret microphones. An instrumentalist is placed in the center, playing a plain low note without strong articulations or modulations, like vibrato or tremolo. One second of quasi-stationary sound was transformed into the spectral domain by discrete Fourier transform (DFT) yielding 128 complex spectra

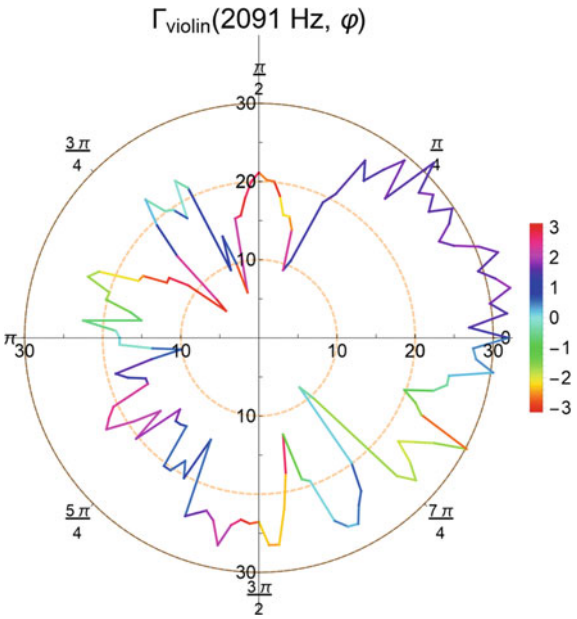

Fig. 11 Measured radiation pattern of one violin frequency

$$P(\boldsymbol{\omega}, \mathbf{r}) = \mathrm{DFT}[p(t, \mathbf{r})] \tag{17}$$

where $\mathbf{r}$ is the position vector of each microphone, consisting of its distance to the origin $\mathbf{r}$ and the angle ϕ between the microphone and the normal vector which is the facing direction of the instrumentalist. Each frequency bin in a complex spectrum has the form $\hat{A}e^{i\varphi}$ with the amplitude $\hat{A}$, the phase φ, Euler's number e and the imaginary unit i. The complex spectra of one violin partial are illustrated in Fig. 11. The amplitude is plotted over the corresponding angle of the microphones, the phase is coded by color. With this setup the radiated sound of 10 instruments has been measured. The investigated instruments are listed in Table 1. Just as in most room acoustical investigations, only partials up to the upper limit of the 8 kHz octave band, i.e. $f_{\max} = 11,314\,\mathrm{kHz}$, are considered. For higher frequencies, the density of partials becomes very high and the signal-to-noise ratio becomes low. Partials are selected manually from the spectrum to find partials, double peaks and to exclude electrical hum etc. reliably.

4.2 The Complex Point Source Model

To compare these musical instruments despite their mostly dissimilar geometries, they are simplified as complex point sources for further investigations. In principle, the complex point source model can be explained easily by Figs. 12 and 13.

Table 1 List of investigated musical instruments and their width at three different distances

Instrument	Width (°)
Accordion	28/19/10
Bagpipe	23/15/8
Crash cymbal	37/25/13
Dizi flute	11/8/4
Double bass	36/24/12
Harmonica	13/9/4
Mandolin	35/24/12
Shakuhachi	11/8/4
Tenor saxophone	11/8/4
Violin	19/13/6

The crash cymbal and the dizi flute have been added after the presentation of preliminary results in Ziemer [50]

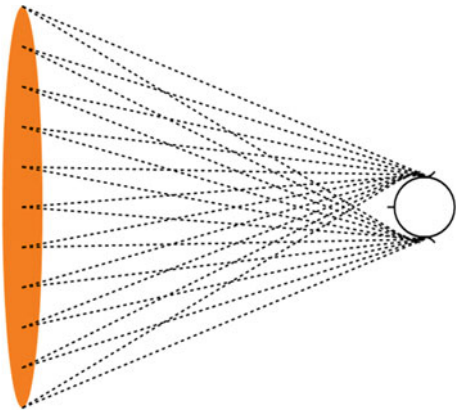

Fig. 12 Schematic sound path from an extended source to the ears. The superposition of radiated sound from all parts of the instrumental body reach both ears

Figure 12 shows a sampled version of the paths that pressure fluctuations undergo from the surface or enclosed air of an extended source to the ears of a listener. Radiations from all parts of the instrument reach both ears. In this consideration we neglect near field effects like evanescent waves and acoustic short circuits. Figure 13 shows a drastic simplification. The instrument is now considered as one point which radiated sound towards all direction, modified by the amplitude and phase that we have measured for the 128 specific angles.

The radial propagation of a point source can be described by the free field Green's function

$$G(r) = \frac{\mathrm{e}^{-ikr}}{r}, \tag{18}$$

where the pressure amplitude decays according to the $1/r$ distance law and the phase shifts according to the wave number $k = 2\pi/\lambda$, where λ is the wave length.

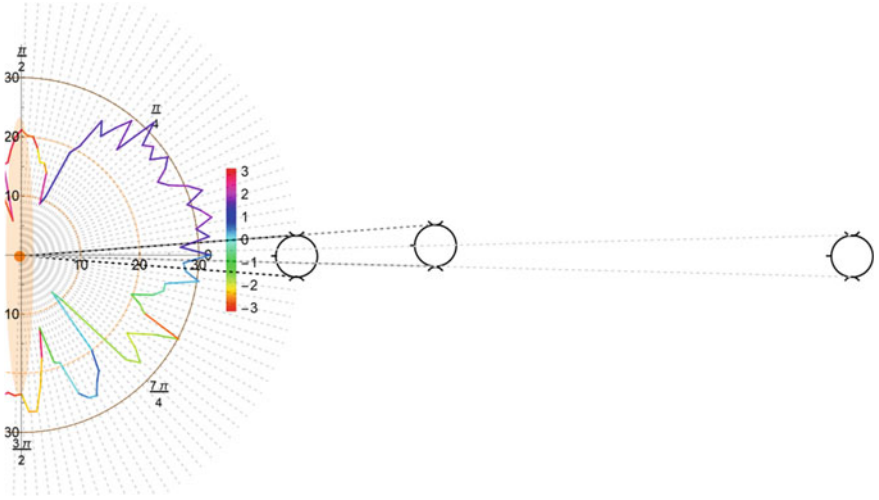

Fig. 13 Ear signals resulting from the complex point source simplification

Covering a circumference with 128 microphones yields one microphone every $\Delta\phi = 2.8°$. The distance between the two ears of a human listener is about 0.15 m. Assuming a listener facing the source point at a distance of 1 m, the distance of the ears correspond to every third microphone, at a distance of 1.5 m every second microphone and at 3 m every microphone. Thus, we can calculate interaural signal differences by comparing every third recording or by propagating all measured signals to a distance of 1.5 and 3 m by Eq. 18 and compare every second or every neighboring propagated microphone recording. This yields a set of $3 \times 128 = 384$ virtual listening positions for which we can calculate ear signals without the use of interpolations.

Neglecting the actual source geometry and considering a musical instrument as a point instead is a rather drastic simplification. Still, the computational benefits are obvious. Furthermore, the model has proven to yield plausible results both physically and perceptually.[35]

4.3 Physical Measures

For all 384 virtual listening positions a number of monaural and binaural physical measures has been calculated. Although no actual listeners are present, the measured and propagated microphone signals are termed "ear signals" in this investigation. Most of them are derived from parameters used in the field of

[35]As has been reported e.g. in Ziemer [49], Ziemer and Bader [52] and Otondo and Rindel [38].

psychoacoustics or room acoustics. But they are adapted to pure, direct, instrumental sound. Due to the vast consensus in the literature,[36] a combination of one monaural and one binaural parameter is searched which best predict the width of musical instruments. The monaural parameter quantifies the strength of bass, the binaural parameter represents the portion of interaural differences compared to interaural coherence. Monaural and binaural parameters are described subsequently.

4.3.1 Monaural Measures

The early low strength $G_{E,\mathrm{low}}$—mentioned in Sect. 2.2, Eq. 5—cannot be applied to pure direct sound as it is the ratio of bass energy in the reverberant field compared to the free field. Therefore, other parameters have been tested, representing the relative strength of low frequencies.

First, all partials f_i below $f_{\max} = 11.314\,\mathrm{kHz}$ are selected manually from the spectrum. As a monaural measure, the fundamental frequency f_1 of each instrumental sound is determined. Likewise, the number of partials I present in the considered frequency region is counted. For harmonic spectra that contain all multiple integers of the fundamental, I should be proportional to $1/f_1$. This is not the case for inharmonic spectra like that of the crash cymbal or instruments like the accordion, which show beatings, i.e. double peaks. Thus, both measures are considered as potential monaural descriptors for a multiple regression analysis. These quantities characterize the source spectrum. They are independent of the listening position.

The amplitude ratio between partials in the 125 and 250 Hz octave bands and in the 500 and 1000 Hz octave bands quantifies bass as a *bass ratio* (BR). A linear and a logarithmic bass ratio

$$\mathrm{BR}_{\mathrm{lin}}(\phi) = \frac{\sum_{f_i \geq 88\,\mathrm{Hz}}^{f_i < 355\,\mathrm{Hz}} \hat{A}^2(f_i)}{\sum_{f_i \geq 355\,\mathrm{Hz}}^{f_i \leq f_{\max}} \hat{A}^2(f_i)} \tag{19}$$

and

$$\mathrm{BR}_{\mathrm{log}}(\phi) = \frac{\sum_{f_i \geq 88\,\mathrm{Hz}}^{f_i < 355\,\mathrm{Hz}} 10\lg\left(\frac{\hat{A}^2(f_i)}{\hat{A}^2(f)_{\min}}\right)}{\sum_{f_i \geq 355\,\mathrm{Hz}}^{f_i \leq f_{\max}} 10\lg\left(\frac{\hat{A}^2(f_i)}{\hat{A}^2(f)_{\min}}\right)} \tag{20}$$

are calculated. Here, $\hat{A}^2(f)_{\min}$ is the lowest amplitude of all partials found in the four octave bands. These two parameters are similar to the *bass ratio* known from room acoustics. In room acoustics, typically reverberation times, early decay times or, sometimes, strength of low frequencies are compared to midrange frequencies.

[36] Refer to the literature cited in Sect. 2.2.

As some instruments create even lower frequencies, and most instruments create much higher frequencies, these two measures can be extended to a relative *bass pressure* (BP) and *bass energy* (BE) in the sound:

$$\mathrm{BP}(\phi) = \frac{\sum_{i=1}^{f_i < 355\,\mathrm{Hz}} \hat{A}(f_i)}{\sum_{f_i \geq 355\,\mathrm{Hz}}^{f_i \leq f_{\max}} \hat{A}(f_i)} \tag{21}$$

$$\mathrm{BE}(\phi) = \frac{\sum_{i=1}^{f_i < 355\,\mathrm{Hz}} \hat{A}^2(f_i)}{\sum_{f_i \geq 355\,\mathrm{Hz}}^{f_i \leq f_{\max}} \hat{A}^2(f_i)} \tag{22}$$

For BP the sum of amplitudes $\hat{A}(f_i)$ of all frequencies below the upper limit of the 250 Hz octave band is compared to the sum of all other considered partials' amplitudes. This value is similar to *BE*, which is the ratio of squared amplitudes. Note that BP^2 does not equal *BE*. If only low-frequency sound is present, all four ratios are undefined as the denominator would be zero. In all other cases they are positive values. The higher the value the higher the sound pressure of the low-frequency components compared to higher partials.

The functions of BE and $\mathrm{BR}_{\mathrm{lin}}$, plotted over the angle, look quite similar. An example is shown in Fig. 14. Especially when transforming the values to a logarithmic scale, BE, $\mathrm{BR}_{\mathrm{lin}}$ and BP look rather similar. This can be seen in Fig. 15, where the logarithm of the three quantities is plotted over angle and scaled to similar magnitudes.

As the monaural parameter is supposed to represent the presence or strength of bass, the *spectral centroid* is a meaningful measure. According to Shimokura et al. [44], *C* is strongly related to the spectral distribution and to W_{IACC}, which had been

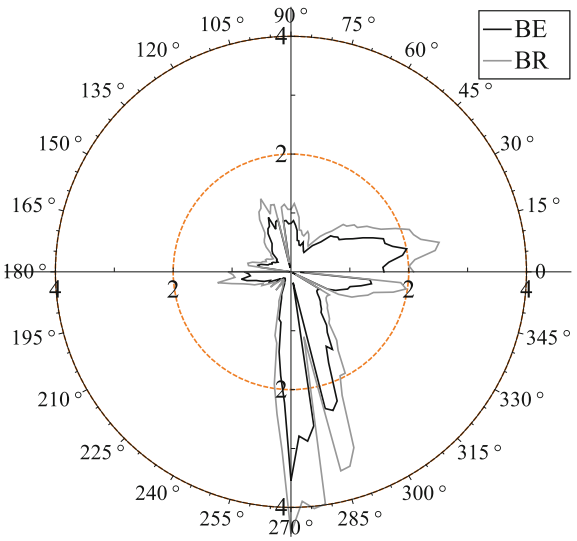

Fig. 14 BE and $\mathrm{BR}_{\mathrm{lin}}$ of a bagpipe, plotted over the listening angle

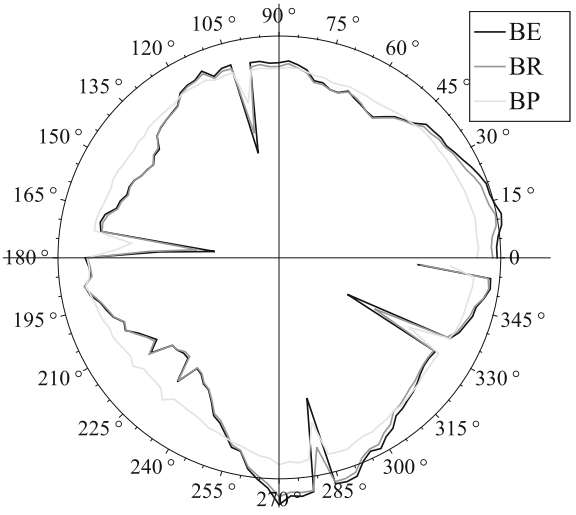

Fig. 15 Logarithmic plot of BE, BR_{lin} and BP of a bagpipe. They are scaled to similar magnitudes

proposed to quantify bass in ASW investigations. Three versions of the spectral centroid are calculated, namely the classic spectral centroid

$$C(\phi) = \frac{\sum_{f=20\,\mathrm{Hz}}^{20\,\mathrm{kHz}} f\hat{A}(f,\phi)}{\sum_{f=20\,\mathrm{Hz}}^{20\,\mathrm{kHz}} \hat{A}(f,\phi)}, \tag{23}$$

where all spectral components are included. The upside of this measure is that even higher partials and noisy components are considered. The downside is that this measure is sensitive to noise of the measurement equipment. This sensitivity is reduced when limiting the bandwidth to the octave bands from 63 Hz to 8 kHz, to get the *band-passed spectral centroid*

$$C_{\mathrm{bp}}(\phi) = \frac{\sum_{f=43\,\mathrm{Hz}}^{11{,}314\,\mathrm{Hz}} f\hat{A}(f,\phi)}{\sum_{f=43\,\mathrm{Hz}}^{11{,}314\,\mathrm{Hz}} \hat{A}(f,\phi)}. \tag{24}$$

The most robust approach is to calculate the spectral centroid only from all manually selected partials

$$C_{\mathrm{part}}(\phi) = \frac{\sum_{i=1}^{I} f_i\hat{A}(f_i,\phi)}{\sum_{i=1}^{I} \hat{A}(f_i,\phi)}. \tag{25}$$

These monaural quantities are independent of the listening distance but they depend on listening angle. Therefore, the mean value over all angles is taken.

In summary, the nine monaural parameters f_1, I, BR_{lin}, BR_{lin}, BP, BE, C, C_{bp} and C_{part} are determined. Monaural measures are independent of the listening

distance whereas source width in degrees is not. Hence, no high correlation between monaural parameters and source extent is expected.

4.3.2 Interaural Measures

As stated before, interaural signal differences are expected to have a larger contribution to width perception than monaural cues. They are calculated from the signals that have been recorded at or propagated to the ear positions of the 384 virtual listeners.

Following the idea of the lateral energy fraction (LEF_{E4}), Eq. 1, the binaural pressure component (BPC) is proposed as the mean ratio between interaural and monaural sound pressure component of all partials

$$\mathrm{BPC}(\mathbf{r}) = \sum_{f_i \geq 88\,\mathrm{Hz}}^{f_i \leq 1{,}414\,\mathrm{Hz}} \frac{|P(f_i, \mathbf{r}_\mathrm{L}) - P(f_i, \mathbf{r}_\mathrm{R})|}{|P(f_i, \mathbf{r}_\mathrm{L}) + P(f_i, \mathbf{r}_\mathrm{R})|} / \mathrm{norm}. \tag{26}$$

for the octave bands from 125 to 1000 Hz. The norm is the bandwidth, i.e. the distance between the actual lowest and highest partial present within these four octave bands. Similarly, the binaural energy component (BEC)

$$\mathrm{BEC}(\mathbf{r}) = \sum_{f_i \geq 88\,\mathrm{Hz}}^{f_i \leq 1{,}414\,\mathrm{Hz}} \frac{(P(f_i, \mathbf{r}_\mathrm{L}) - P(f_i, \mathbf{r}_\mathrm{R}))^2}{(P(f_i, \mathbf{r}_\mathrm{L}) + P(f_i, \mathbf{r}_\mathrm{R}))^2} / \mathrm{norm}. \tag{27}$$

is the ratio between the squared sound pressure difference and the squared sum.

BPC and BEC of a dizi flute are plotted for all listening positions in Figs. 16 and 17. The BPC has higher values, in the BEC some peaks are emphasized compared to the BPC.

It is not meaningful to apply the binaural quality index (BQI), Eq. 4, to the direct instrumental sounds. In room acoustical investigations, the time lag accounts for the fact that lateral reflections might arrive at a listener. These create a maximum interaural time difference of almost ±1 ms. The time lag compensated for this interaural time difference. But under the present free field conditions, all virtual listeners face the source and no reflections occur. Thus, only the interaural correlation coefficient (ICC) is calculated. According to Yanagawa et al. [46], it is the better estimator of ASW, anyway. It equals Eq. 2 if τ is chosen to be 0.1—ICC of a mandolin is plotted in Fig. 18. The same fluctuations as in room acoustical investigations occur.

The interaural difference (IAD), Eq. 6, can be calculated for time windows of 40 ms just as proposed in Griesinger [20]. An example is plotted in Fig. 19. Like C, C_{bp}, and 1—ICC, this measure is sensitive to uncorrelated noise that is present in the recordings.

The ILD and IPD of one partial f_i can easily be calculated by

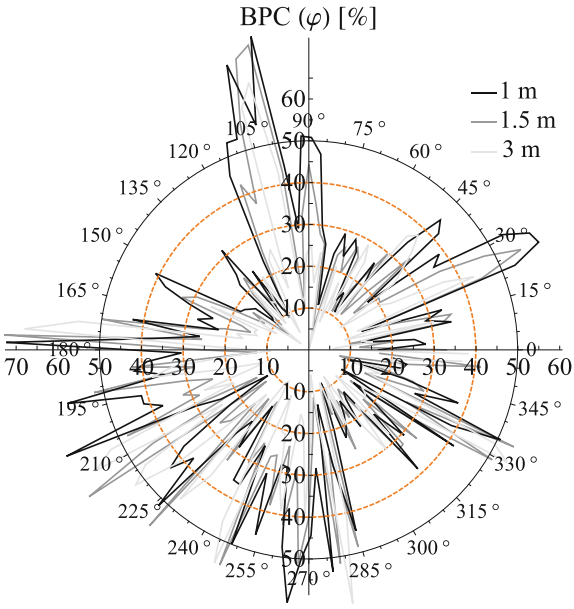

Fig. 16 Binaural pressure component (*BPC*) of a dizi flute at three listening distances plotted over listening angle

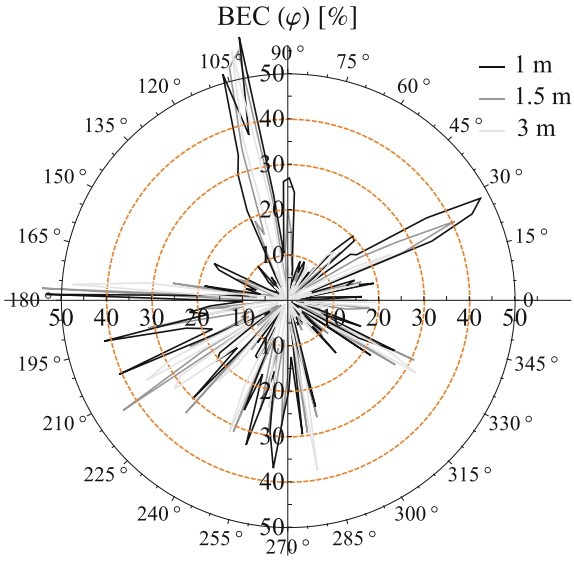

Fig. 17 Binaural energy component (*BEC*) of a dizi flute at three listening distances plotted over listening angle

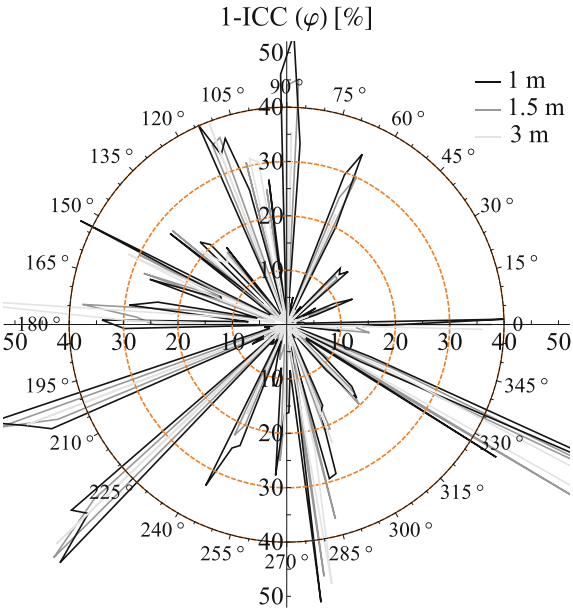

Fig. 18 1—ICC of a mandolin

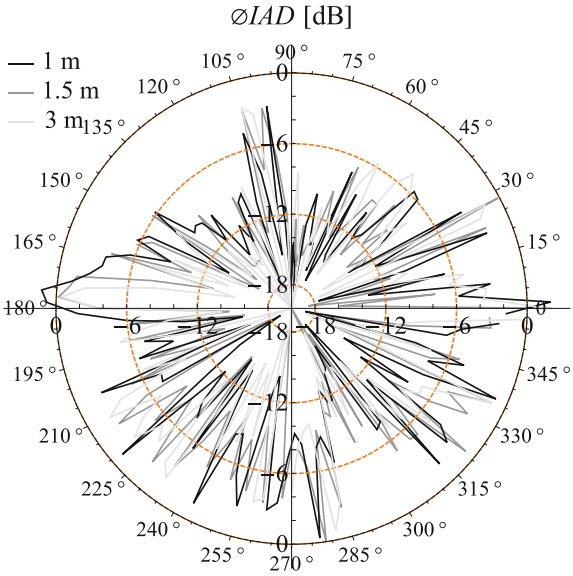

Fig. 19 IAD of a double bass

$$\mathrm{ILD}(f_i, \mathbf{r}) = \left| 20 \lg \left(\frac{\hat{A}(f_i, \mathbf{r}_L)}{\hat{A}(f_i, \mathbf{r}_R)} \right) \right| \tag{28}$$

and

$$\mathrm{IPD}(f_i, \mathbf{r}) = |\varphi(f_i, \mathbf{r}_L) - \varphi(f_i, \mathbf{r}_R)|. \tag{29}$$

Here, $\hat{A}$ is the amplitude and φ the phase. Naturally, the ILD and IPD of loud partials can be heard out more easily by a listener. Thus, they are expected to be more important than those of soft partials. Therefore, they are both weighted by the same factor

$$g(f_i, \mathbf{r}) = \frac{\left| \hat{A}(f_i, \mathbf{r}_L), \hat{A}(f_i, \mathbf{r}_R) \right|_\infty}{\hat{A}(\mathbf{r})_{\max}} \tag{30}$$

which is the larger amplitude of one frequency f_i at both ears L and R, normalized by the highest amplitude of all frequencies at the considered listening position $\hat{A}(\mathbf{r})_{\max}$. The factor g follows the idea of the binaural listening level LL which Ando [2] found to be important for width perception of multi-band noise. Combining Eq. 30 with 28 and 29, respectively, yields the weighted interaural level and phase difference (gILD and gIPD).

To be more close to human perception, the IPD parameter is adjusted by one more step. As mentioned above, the human auditory system is only sensitive to IPD below 1.2 kHz, so only partials below this upper threshold are considered to yield the weighted, band-passed interaural phase difference

$$g\mathrm{IPD}_{\mathrm{bp}}(f_i, \mathbf{r}) = g(f_i, \mathbf{r}) |\varphi(f_i, \mathbf{r}_L) - \varphi(f_i, \mathbf{r}_R)|, f_i \leq 1.2\,\mathrm{kHz}. \tag{31}$$

The evolution from IPD over gIPD to $g\mathrm{IPD}_{\mathrm{bp}}$ can be observed in Figs. 20, 21 and 22. These are plots of a harmonica. The IPD looks somewhat noisy and has two valleys around 20° and 200°. When weighting them with the amplitudes, gIPD looks quite similar. Only the overall magnitudes change. Neglecting all frequencies above 1.2 kHz, the magnitudes are even much lower. Some rather distinct peaks occur at several angles. These coincide with peaks in 1—ICC.

The main difference between the BQI and the $\mathrm{gIPD}_{\mathrm{bp}}$ lies in the fact that the former considers phase inversion not as spatial whereas the latter does. It is emphasized in Damaske and Ando [13] that if the maximum absolute value which determines the BQI comes from a negative value, the listening condition is unnatural.[37] This is evidence that ear signals being in phase and out of phase should be considered as being different in perception.

[37]See Damaske and Ando [13], p. 236.

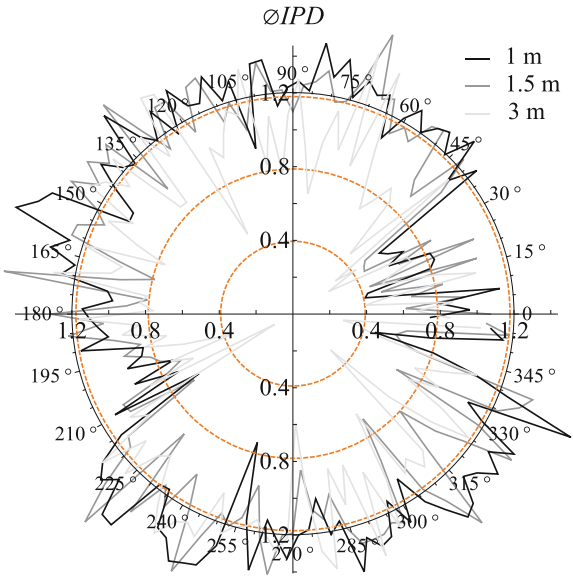

Fig. 20 IPD of the harmonica at all angles and distances

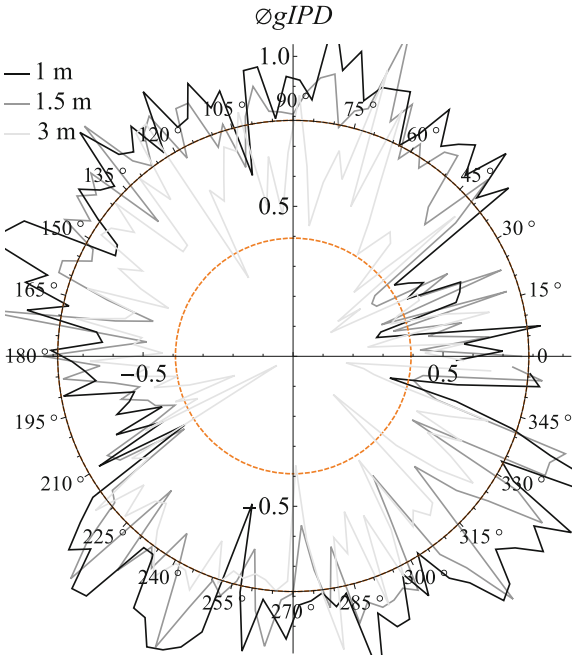

Fig. 21 gIPD of the harmonica at all angles and distances

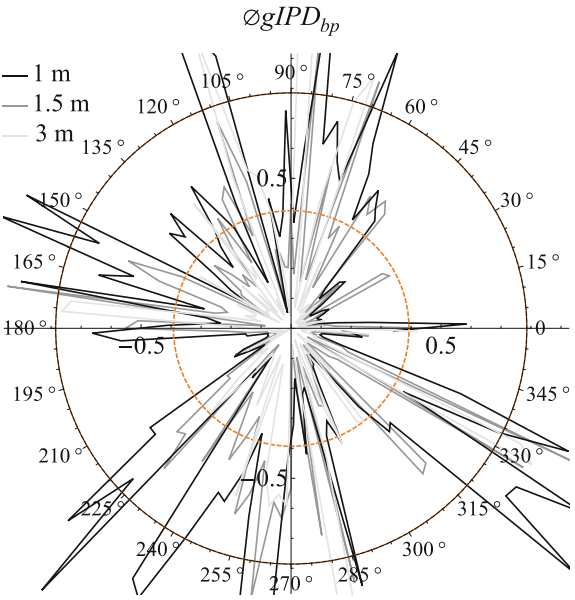

Fig. 22 $g\mathrm{IPD}_{\mathrm{bp}}$ of the harmonica at all angles and distances

In summary, the nine binaural sound field quantities BPC, BEC, 1—ICC, IAD, ILD, IPD, gILD, gIPD and $g\mathrm{IPD}_{\mathrm{bp}}$ are measured. As illustrated in the figures, these measures tend to have lower magnitudes at further distances. This is true for most angles. This behavior is expected, as the source width also decreases with increasing distance. Quantities like RL_{E}, Eq. 8, and $\mathrm{RCC}(t, \tau)$, Eq. 10, are not adopted to the present free field conditions. The first uses delay times of reflections, which are not present in this investigation. The latter assumes that the perceived source extent changes due to the amount and diffusion of reflections. This is not expected for a single note in a free field.

4.4 Results

All sound field quantities that exhibit a significant correlation with source width are listed in Table 2. Here, the Pearson correlation coefficient is listed. The significance level of $p < 0.05$ is indicated by bold numbers, $p < 0.01$ are underlined. Among the monaural measures, the lowest partial f_1, shows a significant negative correlation with width. The number of partials I in the considered frequency region exhibits a highly significant correlation with the source width ($p = 0.001830$). The scatter and the function of the linear regression are plotted in Fig. 23. The width is given in radian. One instrument creates three vertically arranged equidistant points. This is the case because it provides the same I for all three distances. The correlation between $\mathrm{BR}_{\mathrm{log}}$ and width lies slightly above the $p < 0.05$ level ($p = 0.060661$).

Table 2 Pearson correlation for all quantities that exhibit a significant correlation with width

	*f*1	*I*	BRlog	ILD	gILD	gIPDbp	1-CC	BPC	BEC	width
*f*1	1	**−0.676**	−0.135	−0.297	−0.083	0.017	**0.424**	−0.255	−0.342	**−0.396** 0.030350
I	**−0.676**	1	−0.224	**0.543**	0.356	0.051	0.019	**0.481**	**0.546**	**0.545** 0.001830
BRlog	−0.135	−0.224	1	0.048	**−0.409**	−0.208	**−0.704**	−0.047	−0.029	−0.365713 0.060661
ILD	−0.297	**0.543**	0.048	1	**0.620**	**0.442**	0.087	**0.586**	**0.483**	**0.449** 0.012866
gILD	−0.083	0.356	**−0.409**	**0.620**	1	**0.640**	**0.387**	**0.594**	**0.556**	**0.401** 0.028227
gIPDbp	0.017	0.051	−0.208	**0.442**	**0.640**	1	**0.383**	**0.807**	**0.725**	**0.591** 0.000581
1—ICC	**0.424**	0.019	**0.704**	0.087	**0.387**	**0.383**	1	0.248	0.147	**0.401** 0.028211
BPC	0.255	**0.481**	−0.047	**0.586**	**0.594**	**0.807**	0.248	1	**0.972**	**0.654** 0.000087
BEC	−0.342	**0.546**	0.029	**0.483**	**0.556**	**0.725**	0.147	**0.972**	1	**0.646** 0.000114

The significance levels $p < 0.05$ are bold, $p < 0.01$ are underlined. For width, the p-value is given below the correlation coefficient

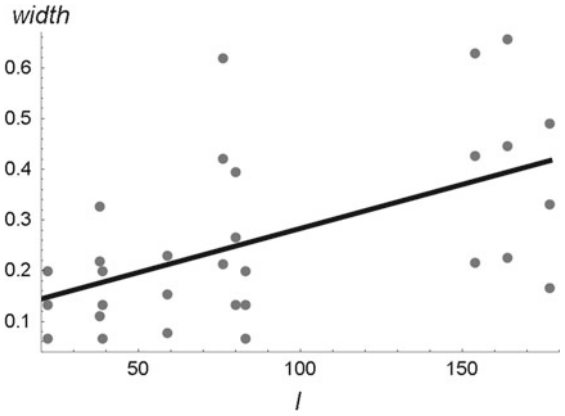

Fig. 23 Source width plotted over the number of partials I (*gray*) and the linear regression function (*black*)

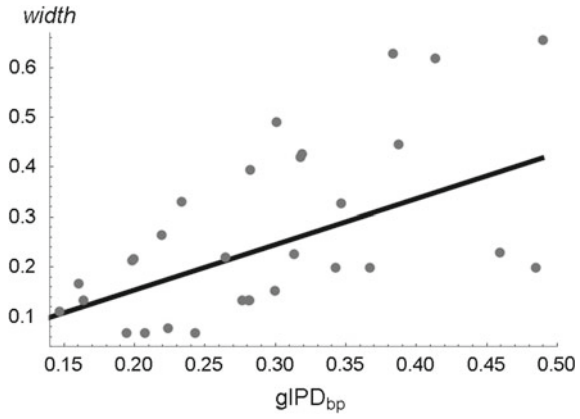

Fig. 24 Source width plotted over $g\mathrm{IPD}_{\mathrm{bp}}$ (*gray*) and the linear regression function (*black*)

As expected, the pair f_1 and I has a highly significant negative correlation. Six of the nine binaural quantities correlate significantly with width. The scatter and the linear regression function of $g\mathrm{IPD}_{\mathrm{bp}}$ are plotted in Fig. 24. 12 of the 15 binaural pairs also correlate significantly with each other, 8 of them on a $p < 0.01$ level. Most important for the multiple regression is the lower left region in the table. A pair of one monaural and one binaural sound field quantity is supposed to explain the source width. 3 monaural and 6 binaural quantities yield 18 potential pairs. However, 6 of them are ineligible, since they exhibit a significant correlation. Thus, they cannot be considered as orthogonal, which is a requirement for a valid multiple linear regression.

Results of multiple regressions with all pairs are summarized in Table 3. All 18 multiple regressions are significant ($p < 0.05$), 14 of them even highly significant ($p < 0.01$). Ineligible pairs that exhibit a correlation with each other are crossed out. Six of the combinations explain over 50 % of the variance, 5 of them are valid pairs. They are highlighted in gray. The linear combination of I and $g\mathrm{IPD}_{\mathrm{bp}}$

Table 3 Explained variance (R^2, top) and significance level (p-value) of multiple regressions between a pair of sound field quantities and source width

	ILD	gILD	gIPDbp	1—ICC	BPC	BEC
f_1	0.277 0.013	0.293 0.009	0.514 0.000058	~~0.55~~ ~~0.000021~~	0.484 0.000131	0.452 0.000295
I	~~0.331~~ ~~0.004~~	0.346 0.003	**0.615** **0.000002**	0.450 0.000315	~~0.497~~ ~~0.000093~~	~~0.471~~ ~~0.000187~~
B_{Rlog}	0.350 0.006	~~0.244~~ ~~0.035~~	0.512 0.000184	~~0.292~~ ~~0.016~~	0.588 0.000024	0.560 0.000052

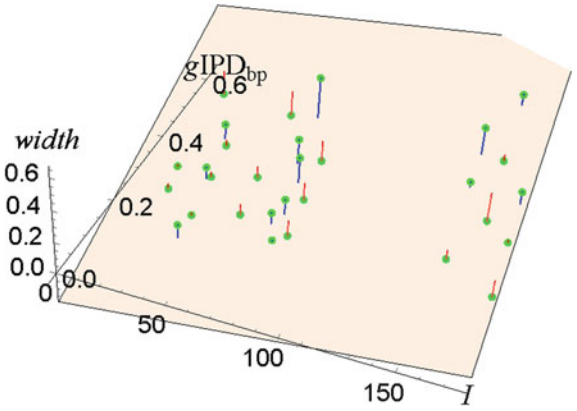

Fig. 25 Source width (*green*) plotted over I and $g\text{IPD}_{bp}$. The actual source width is connected to the predicted width which is based on multiple linear regression (transparent plane)

explains R^2 = 61.5 % ($p = 0.000002$) of the variance of source width. At an earlier state of research, R^2, the coefficient of determination, was 56 % ($p = 0.001601$) when considering only 8 instruments (Ziemer [50]). With a larger sample, including one inharmonic instrument, the results of the multiple linear regression improved. The result is illustrated in Fig. 25. Over-estimated widths are connected to the prediction plane with red lines, under-estimated widths with blue lines. It can be seen that the multiple linear regression yields a fair prediction of source width. This is even true for the extremes. No drastic outliers can be observed.

Some nonlinear combinations of I and $g\text{IPD}_{bp}$ yield slight improvements of the regression. Using the logarithm of the two, R^2 = 63.1 % of the variance is predictable, using their square root, R^2 becomes 63.2 %. A more effective nonlinear combination is similar to Eq. 7 as proposed by Ando [2], like

$$\text{ASW}_{pre} = aI^{1/3} + bg\text{IPD}_{bp}^{2/3} + c \tag{32}$$

which explained R^2 = 63.4 % of the variance.

5 Discussion

In this investigation, the radiation characteristics of 10 musical instruments has been measured. The radiated sound field is either directly measured at or propagated to 384 listening positions. Here, quantities from the field of psychoacoustics and subjective room acoustics have been calculated. Based on a pair of one monaural and one binaural parameter, the actual source width could be predicted with a fair precision. The best monaural predictor was the plain number of partials I in the considered frequency range. It is an even better predictor than the fundamental frequency or several measures of bass energy. Although the binaural pressure and energy components BPC and BEC exhibited a higher correlation with source extent, and even with a lower p-value, the weighted interaural phase difference below 1.2 kHz $g\mathrm{IPD}_{\mathrm{bp}}$ turned out to be the best predictor of source width, in combination with I.

This means that the number of partials might play a role in width perception. On the one hand, I is related to bass strength. The lower the fundamental frequency of musical instruments, the more partials in the spectrum tend to have an audible amplitude. From the literature, bass strength is already known to be related to the perception of source width. On the other hand, I is also closely related to spectral density. Spectral density might also be related to source extent and affect the perception of width.

Both versions of ILD significantly correlated with source width. This is in good agreement with the results derived from Potard and Burnett [39], that ILD are important for the recognition of shapes. It also seems to confirm the finding by Schroeder [43] that ILD are an important factor for a spatial sound impression. But $g\mathrm{IPD}_{\mathrm{bp}}$ gave the better prediction of width. This might imply that phase difference is an even more important parameter than level difference. This might be true in both a technical and a perceptual sense. It is interesting to see that a psychoacoustically motivated modification distinctly improved the results. A significant relationship could neither been found for IPD and width ($p = 0.289090$) nor between gIPD and width ($p = 0.114490$). But when considering only phase differences below the threshold of IPD perception, a high significance level is reached. This could mean that lower frequencies give more reliable cues for width perception. Of course, there are additional physical aspects: Considering a musical instrument as complex point source is a drastic simplification which is meaningful for low frequencies but it does not reflect the actual radiation characteristics of high frequencies well. Furthermore, due to the large wavelengths of low frequencies, slight misplacements of microphones hardly affect their measured phase. But for high frequencies, small misplacements can result in larger phase errors. As most of the considered partials lie above 1.2 kHz, the filtering eliminates these phase errors.

On the one hand, explaining 61.5 % of the variance is not very much. On the other hand, the number of considered instruments and listening distances is rather low. A higher R^2 is expected for a larger data set. This has proven to be true already: In an earlier state of this investigation, when only 8 instruments had been

measured, R^2 was 56 %. As even subjective judgments about perceived width provide a high variance, R^2 = 61.5 % might be sufficient for many applications. Considering and controlling the interaural phase differences of loud frequencies as well as the number of partials might be the right way to analyze and manipulate perceived source width. Of course, ICLDs and ICPDs in a stereo or surround setup do not create the same ILDs and IPDs. Zotter and Frank [54] have demonstrated that ICCC and IACC are proportional within a certain range. Naturally, ILD and IPD are lower than ICLD and ICPD. However, for a sweet spot, a simplified HRTF as proposed in Kling and Riggs [29] (p. 351) or a publicly available HRTF as published e.g. in Blauert et al. [7] and Algazi et al. [1] can be used to translate inter-channel differences to inter aural differences. In ambisonics and wave field synthesis systems where several listeners can move through an extended listening area, another method is necessary. One solution is to sample the listening area into a finite number of potential listening positions and create the desired $g\mathrm{IPD}_{\mathrm{bp}}$ here. This could be achieved by means of a high-order point multipole source as implemented in Corteel [12]. Alternatively, a rather coherent localization signal at each note onset is followed by the desired $g\mathrm{IPD}_{\mathrm{bp}}$ similar to the approach of Ziemer and Bader [53]. Likewise, DirAC encoding follows the idea to give one localization cue and one width cue. Such a coding could be used to give source position and $g\mathrm{IPD}_{\mathrm{bp}}$ as metadata.

6 Prospects

A reliable knowledge about the auditory perception of source width and the sound field at the listeners' ears is a powerful foundation for many applications. It could act as the basis of audio monitoring tools in recording studios to display perceived source width instead of plain channel correlations. This helps music producers to achieve the desired spatial impression. For channel-based audio systems, control over interaural cues is possible for a sweet spot if the loudspeaker positions are fixed and a HRTF is implemented. When using object-based audio coding, the desired interaural sound field quantities can be stored as metadata. This way, the approach can be adopted for a flexible use with arbitrary loudspeaker constellations. Instrument builders could focus on manipulating $g\mathrm{IPD}_{\mathrm{bp}}$ in a preferred listening region to achieve the desired perceived source extent. For example, the right radiation pattern could make a source sound narrow at one angle and more broad at another angle. Musical instruments for practicing could be designed to create a wider sound impression for the instrumentalist for a greater sound enjoyment. Then, instruments for performance create this sound impression for the audience. Simple measurement tools or advanced physical modeling software could support the work of instrument builders. Room auralization software can sound more realistic if it focuses on calculating the relevant parameters with high precision. Implementing radiation patterns of extended sources on sound field synthesis technologies, like

higher order ambisonics and wave front synthesis, can make the sound broader and more realistic. When concentrating on $g\mathrm{IPD}_{\mathrm{bp}}$ of partials as perceptually relevant parameters, computation time can be saved by synthesizing these cues instead of the whole radiation characteristics or other irrelevant parameters. This is again interesting for advancements in electric and electronic instruments. Electric pianos could sound more realistic, if the right auditory cues are recreated which make an actual grand piano sound this broad. Electric guitars could be widened and narrowed by turning one knob on the guitar amps which creates the desired monaural and interaural cues for a sweet spot or a limited listening region.

Until now, the presented approach lacks psychoacoustic proof. Listening tests under controlled conditions can bring reliable results concerning the relationship between sound radiation characteristics and perceived source extent. A prediction of source width may be more precise and especially more close to human perception when auditory processing is considered. Implementing binaural loudness and masking algorithms or even higher states of auditory processing is very promising to explain perceived source width in more detail.

References

1. Algazi, V.R., Duda, R.O., Thompson, D.M., Avendano, C.: The CIPIC HRTF database. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, NY, pp. 99–102 (2001)
2. Ando, Y.: Auditory and Visual Sensation. Springer, New York (2010)
3. Baalman, M.: On Wave Field Synthesis and Electro-acoustic Music, with a Particular Focus on the Reproduction of Arbitrarily Shaped Sound Sources. VDM, Saarbrücken (2008)
4. Beranek, L.L.: Concert Halls and Opera Houses: Music, Acoustics, and Architecture, 2nd edn. Springer, New York (2004)
5. Blau, M.: Correlation of apparent source width with objective measures in synthetic sound fields. Acta Acust. United Acust **90**(4), 720–730 (2004)
6. Blauert, J.: Spatial Hearing. The Psychophysics of Human Sound Source Localization (Revised edn.). MIT Press, Cambridge (1997)
7. Blauert, J., Brüggen, M., Hartung, K., Bronkhorst, A.W., Drullmann, R., Reynaud, G., Pellieux, L., Krebber, W., Sottek, R.: The AUDIS catalog of human HRTFs. In: Proceedings of the 16th International Congress on Acoustics, vol. 4, pp. 2901–2902, Seattle (1998)
8. Blauert, J., Cobben, W.: Some consideration of binaural cross correlation analysis. Acta Acust. United Acust **39**(2), 96–104 (1978)
9. Blauert, J., Lindemann, W.: Auditory spaciousness: some further psychoacoustic analyses. J. Acoust. Soc. Am. **80**(2), 533–542 (1986)
10. Bradley, J.S., Reich, R.D., Norcross, S.G.: On the combined effects of early- and late-arriving sound on spatial impression in concert halls. J. Acoust. Soc. Am. **108**(2), 651–661 (2000)
11. Cabrera, A.: Pseudo-stereo techniques. CSound implementations. Csound J. **14** (Article number 3) (2011)
12. Corteel, E.: Synthesis of directional sources using wave field synthesis, possibilities, and limitations. EURASIP J. Adv. Sign. Process. Article ID 90509 (2007)
13. Damaske, P., Ando, Y.: Interaural crosscorrelation for multichannel loudspeaker reproduction. Acta Acust. United Acust **27**(4), 232–238 (1972)

14. de Vries, D., Hulsebos, E.M., Baan, J.: Spatial fluctuations in measures for spaciousness. J. Acoust. Soc. Am. **110**(2), 947–954 (2001)
15. Deutsches Institut für Normung.: Akustik — Messung von Parametern der Raumakustik — Teil 1. Aufführungsräume (ISO 3382-1:2009); Deutsche Fassung EN ISO 3382-1:2009 (2009)
16. Faller, C.: Pseudostereophony revisited. In: 118th Audio Engineering Society Convention, Barcelona (2005)
17. Friedrich, H.J.: Tontechnik für Mediengestalter. Töne hören — Technik verstehen — Medien gestalten. Springer, Berlin (2008)
18. Gade, A.C.: Acoustics in halls for speech and music. In: Rossing, T.D. (ed.) Handbook of Acoustics, Chapter 9, pp. 301–350. Springer, Berlin (2007)
19. Gerzon, M.A.: The design of precisely coincident microphone arrays for stereo and surround sound. In: 50th Audio Engineering Society Convention, London (1975)
20. Griesinger, D.: Objective measures of spaciousness and envelopment. In: AES 16th International Conference: Spatial Sound Reproduction, Rovaniemi (1999)
21. Haas, H.: Einfluss eines Einfachechos auf die Hörsamkeit von Sprache. Acustica **1**, 49–58 (1951)
22. Hamidovic, E.: The Systematic Mixing Guide. Systematic Productions, Melbourne (2012)
23. Heller, A.J.: Is my decoder ambisonic? In: 125th Audio Engineering Society Convention, San Francisco, CA (2008)
24. Hirvonen, T., Pulkki, V.: Center and spatial extent of auditory events as caused by multiple sound sources in frequency-dependent directions. Acta Acust. United Acust. **92**(2), 320–330 (2006)
25. Jacques, R., Albrecht, B., Melchior, F., de Vries, D.: An approach for multichannel recording and reproduction of a sound source directivity. In: 119th Audio Engineering Society Convention, New York (2005)
26. Kaiser, C.: 1001 Mixing Tipps. mitp, Heidelberg (2012a)
27. Kaiser, C.: 1001 Recording Tipps. mitp, Heidelberg (2012b)
28. Kaiser, C.: 1001 Mastering Tipps. mitp, Heidelberg (2013)
29. Kling, J.W., Riggs, L.A. (eds.): Woodworth & Schlossberg's Experimental Psychology, 3rd edn. Holt, Rinehart and Winston, New York (1971)
30. Laitinen, M.-V., Philajamäki, T., Erkut, C., Pulkki, V.: Parametric time-frequency representation of spatial sound in virtual worlds. ACM Trans. Appl. Percept. **9**(2) (2012)
31. Laitinen, M.-V., Walther, A., Plogsties, J., Pulkki, V.: Auditory distance rendering using a standard 5.1 loudspeaker layout. In: 139th Audio Engineering Society Convention, New York, NY (2015)
32. Levinit, D.J.: Instrument (and vocal) recording tips and tricks. In: Greenbaum, K., Barzel, R. (eds.) Audio Anecdotes, vol. I, pp. 147–158. A K Peters, Natick (2004)
33. Lindemann, W.: Extension of a binaural cross-correlation model by contralateral inhibition. ii. the law of the first wave front. J. Acoust. Soc. Am. **80**(6), 1623–1630 (1986)
34. Maempel, H.-J. (2008). Medien und Klangästhetik. In: Bruhn, H., Kopiez, R., Lehmann, A.C. (eds.) Musikpsychologie. Das neue Handbuch, pp. 231–252. Rowohlt, Reinbek bei Hamburg (2008)
35. Martín, R.S., Witew, I.B., Arana, M., Vorländer, M.: Influence of the source orientation on the measurement of acoustic parameters. Acta Acust. United Acust. **93**(3), 387–397 (2007)
36. Mason, R., Brookes, T., Rumsey, F.: The effect of various source signal properties on measurements of the interaural crosscorrelation coefficient. Acoust. Sci. Technol. **26**(2), 102–113 (2005)
37. Okano, T., Beranek, L.L., Hidaka, T.: Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (ASW) in concert halls. J. Acoust. Soc. Am. **104**(1), 255–265 (1998)
38. Otondo, F., Rindel, J.H.: The influence of the directivity of musical instrument in a room. Acta Acust. United Acust **90**, 1178–1184 (2004)
39. Potard, G., Burnett, I.: A study on sound source apparent source shape and wideness. In: Proceedings of the 2003 International Conference on Auditory Display, Boston, MA (2003)

40. Potard, G., Burnett, I.: Decorrelation techniques for the rendering of apparent sound source width in 3d audio displays. In: Proceedings of the 7th International Conference of Digital Audio Effects, Naples (2004)
41. Rogers, S.E.: The art and craft of song mixing. In: Greenbaum, K., Barzel, R. (eds.) Audio Anecdotes, vol. II, pp. 29–38. A K Peters, Natick (2004)
42. Ross, B., Tremblay, K.L., Picton, T.W.: Physiological detection of interaural phase differences. J. Acoust. Soc. Am. **121**(2), 1017–1027 (2007)
43. Schroeder, M.R.: An artificial stereophonic effect obtained from using a single signal. In: 9th Audio Engineering Society Convention, New York, NY (1957)
44. Shimokura, R., Tronchin, L., Cocchi, A., Soeta, Y.: Subjective diffuseness of music signals convolved with binaural impulse responses. J. Sound Vibr. **330**, 3526–3537 (2011)
45. Slavik, K.M., Weinzierl, S.: Wiedergabeverfahren. In: Weinzierl, S. (ed.) Handbuch der Audiotechnik, Chapter 11, pp. 609–686. Springer, Berlin (2008)
46. Yanagawa, H., Anazawa, T., Itow, T.: Interaural correlation coefficients and their relation to the perception of subjective diffuseness. Acta Acust. United Acust. **71**(3), 230–232 (1990)
47. Yanagawa, H., Tohyama, M.: Sound image broadening by a single reflection considering temporal change of interaural cross-correlation. Acta Acust. United Acust. **87**(2), 247–252 (2001)
48. Yanagawa, H., Yamasaki, Y., Itow, T.: Effect of transient signal length on cross-correlation functions in a room. J. Acoust. Soc. Am. **84**(5), 1728–1733 (1988)
49. Ziemer, T.: Sound radiation characteristics of a shakuhachi with different playing techniques. In: Proceedings of the International Symposium on Musical Acoustics, Le Mans, pp. 549–555 (2014)
50. Ziemer, T.: Adapting room acoustic parameters to explain apparent source width of direct sound. In: 'Musik und Wohlbefinden'. 31. Jahrestagung der DGM, Oldenburg, pp. 40–41 (2015)
51. Ziemer, T.: Wave field synthesis. In: Handbook of Systematic Musicology. Springer, Berlin (in Print) (2016)
52. Ziemer, T., Bader, R.: Complex point source model to calculate the sound field radiated from musical instruments. In: Proceedings of Meetings on Acoustics, vol. 25 (2015a)
53. Ziemer, T., Bader, R.: Implementing the radiation characteristics of musical instruments in a psychoacoustic sound field synthesis system. In: 139th Audio Engineering Society Convention, New York, NY (2015b)
54. Zotter, F., Frank, M.: Efficient phantom source widening. Arch. Acoust. **38**(1), 27–37 (2013)
55. Zotter, F., Frank, M., Kronlachner, M., Choi, J.-W.: Efficient phantom source widening and diffuseness in ambisonics. In: Proceedings of the EAA Joint Symposium on Auralization and Ambisonics, Berlin (2014)
56. Zwicker, E., Fastl, H.: Psychoacoustics. Facts and Models (Second updated edn.). Springer, Berlin (1999)

Author Biography

Dr. Tim Ziemer is a musicologist, mainly working in the field of applied psychoacoustics from wave field synthesis over music information retrieval to instrument acoustics. His research interests include the perception of spaciousness and its relation to sound radiation characteristics of musical instruments. Tim Ziemer has worked with research teams at the University of Hamburg and the National Institute of Informatics Tokyo. He is a freelance author for a renowned computer magazine and has a professional background in teaching, music production, room and building acoustics, concert logistics and cultural administration.