PAN
ARCHIVES OF ACOUSTICS
Vol. **38**, No. 3, pp. 311–319 (2013)

DOI: 10.2478/aoa-2013-0038

Source Width of Frontal Phantom Sources: Perception, Measurement, and Modeling

Matthias FRANK

Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz
Inffeldgasse 10/3, 8010 Graz, Austria; e-mail: frank@iem.at

(*received July 4, 2012; accepted February 8, 2013*)

Phantom sources are known to be perceived similar to real sound sources but with some differences. One of the differences is an increase of the perceived source width. This article discusses the perception, measurement, and modeling of source width for frontal phantom sources with different symmetrical arrangements of up to three active loudspeakers. The perceived source width is evaluated on the basis of a listening test. The test results are compared to technical measures that are applied in room acoustics: the inter-aural cross correlation coefficient (IACC) and the lateral energy fraction (LF). Adaptation of the latter measure makes it possible to predict the results by considering simultaneous sound incidence. Finally, a simple model is presented for the prediction of the perceived source width that does not require acoustic measurements as it is solely based on the loudspeaker directions and gains.

Keywords: source width, phantom source, stereophony, IACC, LF, energy vector.

1. Introduction

It is known from concert hall acoustics that reflections from walls and the ceiling influence the perceived spatial extent of instruments. This perceived size of a sound source is often called *auditory* or *apparent source width* (ASW). There are technical measures that correlate with the perception of ASW (Schroeder *et al.*, 1974; Barron, Marshall, 1981; Blauert, Lindemann, 1986a; Hidaka *et al.*, 1995; Morimoto, Iida, 2005). The two most common measures are the inter-aural cross correlation coefficient (IACC) using a dummy head and the lateral energy fraction (LF) using a combination of an omni-directional and a figure-of-eight microphone.

A *phantom source* is an auditory event that is perceived at a location where there is no real source present (Wendt, 1963), i.e. a location between loudspeakers. The simplest method to create a phantom source with two loudspeakers is *stereophony* (Blumlein, 1958) which uses level and/or time-delay differences between the loudspeakers to control the location of the phantom source. Other methods, such as *multiple-direction amplitude panning* (Pulkki, 1999) or *Ambisonics* (Daniel, 2001), use more than two loudspeakers for the creation of a single phantom source. It is known that phantom sources differ from real sources with respect to the perceived source width. Different terms have been used to describe this difference: image focus (Martin *et al.*, 1999), locatedness (Simon *et al.*, 2009), or spatial spread (Pulkki, 1999).

Studies show that the loudspeaker spacing influences the perceived source width (Frank *et al.*, 2011; Kin, Plaskota, 2011). However, there is no systematic study about how the number of loudspeakers and their directions and gains are related to the width of phantom sources. To start discussion about this relationship, this article presents a study with loudspeaker arrangements of various width in a listening setup with dominant direct sound. The loudspeaker arrangements are symmetrical to the 0° axis and use one, two, or three loudspeakers that play the same signal.

The second section of this paper describes a listening test that evaluates the perceived source width. The impulse responses of the loudspeakers in the listening test setup were also measured with a dummy head and a microphone array. From these measurements, section three derives the IACC and LF measures. These measures are compared to the results of the listening test. The third section also presents an adaptation of LF to make it a valid predictor for the experimental data. Section four introduces a simple model for the perceived source width that predicts the data without acoustic measurement, just based on the gains and directions of the loudspeakers and assuming no dependency on room reflections.

In this paper, the directions of L loudspeakers are vectors of unit length $\boldsymbol{\theta}_l = [\cos(\phi_l),\ \sin(\phi_l)]^{\mathrm{T}}$ that depend on their azimuth angle ϕ_l in the horizontal plane, see Fig. 1. For each loudspeaker $l \in \{1, \ldots, \mathrm{L}\}$, the scalar weight g_l denotes its adjustable gain. Impulse responses are represented by $h(t)$.

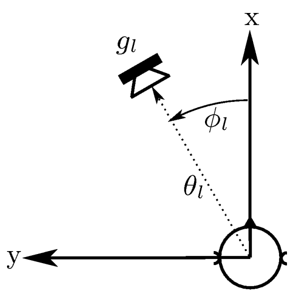

Fig. 1. Reference system used in this paper.

2. Perceptual evaluation

This section describes the method and results of the perceptual evaluation of phantom source width by a listening test. The goal of the measurements and modeling presented later is to find valid technical estimators for the data of the time consuming listening test.

2.1. Method

The evaluation studied perception of source width for one, two, and three active loudspeakers playing the same signal and different spacings between them. Figure 2 shows the test setup with 17 Genelec 8020 loudspeakers: loudspeaker 0 at 0°, and loudspeaker pairs $1 \ldots 8$ at $\pm 5° \ldots \pm 40°$. The setup was built in the center of a 11 m × 11 m × 5 m room with a reverberation time within the limits of ITU-R BS.1116-1 (ITU, 1997). The distance to the central listening position was 2.5 m for each loudspeaker, which lies within the effective critical distance. The height of the loudspeakers (referred to halfway between woofer and tweeter) was adjusted to 1.2 m, which was also the ear height of the subjects. The control of the entire listening test, as well as the creation of the loudspeaker signals used the open source software pure data (freely available on http://puredata.info/downloads) on a PC with an RME HDSPe MADI and RME M-16 DA D/A converters. All measurements in the remainder of this paper were using the same setup and conditions.

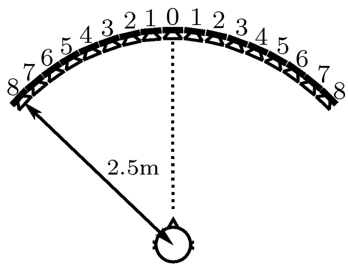

Fig. 2. Loudspeaker setup used for the perceptual evaluation and measurements.

Table 1 shows the angles and gains of the loudspeakers for each of the 16 conditions. Empty entries in the table mean that the corresponding loudspeakers

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

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

are not active. Condition 0 is a single loudspeaker playing from 0° in front of the listener. Conditions 10 ... 80 correspond to 2-channel stereophony on frontally centered pairs of loudspeakers with the same amplitude gain and aperture angles ranging from 10° to 80°. In conditions $C10 \ldots C80$ the central loudspeaker (loudspeaker 0 at 0°) is added with the same gain. These additional conditions aim at extending the applicability of the relationships obtained to arbitrary amplitude panning methods that use more than two loudspeakers, such as multiple-direction amplitude panning (Pulkki, 1999) or Ambisonics (Daniel, 2001). Note that condition $C60$ has not been tested due to an error in the playback software. The gains in all conditions were normalized to a constant overall energy which results in gains that depend on the number of active loudspeakers. Furthermore, the symmetrical arrangement aims at creating a phantom source direction of 0° for all conditions in order to exclude differences in the localization direction between the conditions.

Fourteen subjects participated in the listening test; their individual hearing loss was less than 12 dB between 250 Hz and 8 kHz. All of them were members of a trained expert listening panel (Sontacchi *et al.*, 2009; Frank *et al.*, 2010; Frank, Sontacchi, 2012) and had already participated in listening tests about source width (Frank *et al.*, 2011; Zotter *et al.*, 2011). Each of the 16 conditions was presented seven times for each subject in random order. The stimulus was 1.5 s of pink noise at a level of 65 dB(A). The subjects were allowed to repeat the stimulus at will by pressing a button on a keyboard. They were asked to measure the perceived source width in terms of an index and to write their answer in a questionnaire. The index expressed the perceived width in terms of numbers on the increasingly wide, nested loudspeaker pairs according to Fig. 2. It was also possible to use half indices to rate perceived widths that are between adjacent indices which results in a possible resolution of 5°. The subjects were told to face forward but there was no head fixation in order to allow small head movements. It has been shown that these movements are performed unconsciously (Blauert, 1983) but are important for localization (Mackensen, 2008) and assessment of spatial impression (Brookes *et al.*, 2007).

2.2. Results

The answers were averaged over all subjects and all repetitions. No subjects were excluded from the results. Figure 3 shows the resulting mean value and corresponding 95% confidence interval of the perceived source width for each condition. The results in the figure were arranged to ascending aperture angle between the two outmost active loudspeakers.

Within the 2-channel conditions 10 ... 80, an increase of the aperture angle yields an increase of the

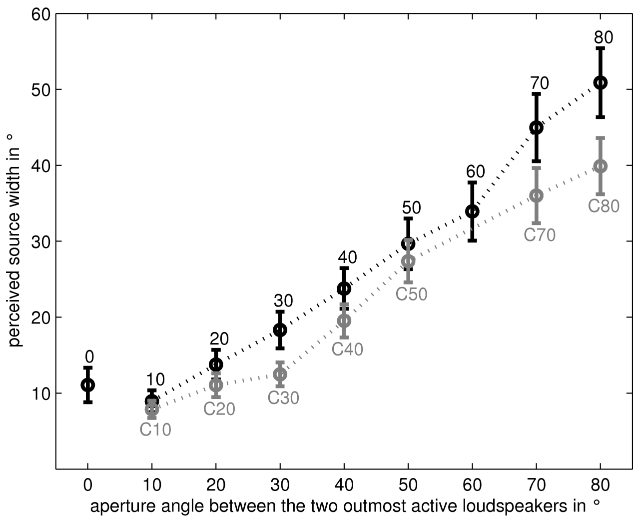

Fig. 3. Mean and 95% confidence interval of the perceived source width for each condition, arranged to ascending aperture angle between the two outmost active loudspeakers.

perceived source width. An analysis of variance confirms the aperture angle as a highly significant factor ($p \ll 0.001$). This holds true for the conditions with the additional central loudspeaker $C10 \ldots C80$. Comparing both groups with and without the central loudspeaker, the addition of the central loudspeaker yields a highly significant decrease of the perceived source width. This relation agrees with the findings in (Kin, Plaskota, 2011) and can be explained by the fact that the active central loudspeaker decreases the relative share of lateral sound. Note that in pairwise comparisons of the corresponding conditions with and without central loudspeaker, differences between the mean values of $C10/10$ and $C50/50$ are not significant ($p > 0.05$).

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

3. Technical measures

This section presents technical measures obtained by acoustic measurements and their correlation to the listening test results above. As these measures are typically applied to reverberant concert halls, this section examines the suitability of the measures for the prediction of phantom source width. The measurements were performed on exactly the same experimental setup and environment as in the listening test.

3.1. Inter-Aural Cross Correlation Coefficient (IACC)

In order to calculate the inter-aural cross correlation coefficient (IACC), binaural impulse responses were measured for each loudspeaker using a B&K 4128C dummy head. For each condition, h_{left} and h_{right} are the impulse responses of the left and the right ear of the dummy head, respectively. They are calculated as the linear superposition of the binaural impulse responses for each loudspeaker $h_{l,\mathrm{left}}$ and $h_{l,\mathrm{right}}$ with the appropriate loudspeaker gains g_l according to Table 1

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

The IACC is defined as the maximum of the inter-aural cross correlation function (IACF), cf. (ISO, 2009)

$$\mathrm{IACF}(\tau) = \frac{\int\limits_{t_1}^{t_2} h_{\mathrm{left}}(t) h_{\mathrm{right}}(t+\tau)\,\mathrm{d}t}{\sqrt{\left[\int\limits_{t_1}^{t_2} h^2_{\mathrm{left}}(t)\,\mathrm{d}t\right]\left[\int\limits_{t_1}^{t_2} h^2_{\mathrm{right}}(t)\,\mathrm{d}t\right]}}, \tag{2}$$

$$\mathrm{IACC} = \max\nolimits_{\tau\in[-1\mathrm{ms};1\mathrm{ms}]} |\mathrm{IACF}(\tau)|, \tag{3}$$

Typically, the observation time is set to the first 80 ms of the impulse responses, i.e. $t_1 = 0$ ms and $t_2 = 80$ ms. As this version of the IACC considers only the early part of the impulse responses, it is called early IACC or $\mathrm{IACC_E}$. Furthermore, the IACC is mostly not calculated for the broadband signals, but separately for three octave bands around 500 Hz, 1 kHz, and 2 kHz (Hidaka *et al.*, 1995). The three correlation coefficients are averaged. Here the early IACC is employed for the three octave bands, denoted as $\mathrm{IACC_{E3}}$.

There are different values for the perceptually just noticeable difference (JND) of $\mathrm{IACC_{E3}}$ in the literature: 0.075 (ISO, 2009; Cox *et al.*, 1993), 0.05–0.08 (Okano, 2002), 0.038 (Blau, 2002).

Figure 4 draws the computed $1-\mathrm{IACC_{E3}}$ measures of all conditions in relation to the mean of the perceived source width from the listening test. The $\mathrm{IACC_{E3}}$ values range from 0.8 to 0.95 and cover 2–4 JNDs which predicts a poor discriminability that is in contradiction to the results of the listening test results, cf. Fig. 3. The value of $\mathrm{R}^2 = 0.65$ for the coefficient of determination reveals a fair correlation between the listening test results and the objective measure. Altogether, $\mathrm{IACC_{E3}}$ does not seem to be an optimal predictor of the perceived source width in case of simultaneous sound incidence. It obviously refers to a longer temporal structure of the impulse responses. In other cases, the $\mathrm{IACC_{E3}}$ is a better predictor if the temporal structure of the loudspeaker signals is manipulated, e.g., by decorrelation algorithms (Zotter *et al.*, 2011), which is not the case in the experimental conditions, here.

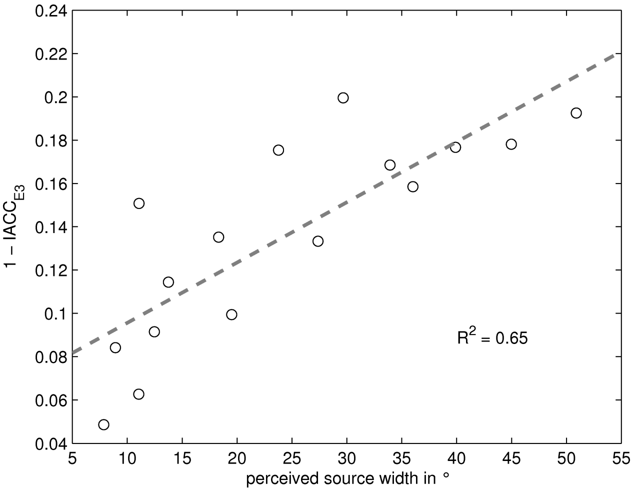

Fig. 4. Regression of $1-\mathrm{IACC_{E3}}$ to the listening test results.

3.2. Lateral Energy Fraction (LF)

The lateral energy fraction (LF) is also used to describe width. It is derived from the impulse response measurements using an omni-directional microphone and a figure-of-eight microphone, yielding the responses $h_\circ$ and h_∞, respectively. For each condition, both responses are computed from the linear superposition of the individual impulse responses of each loudspeaker $h_{l,\circ}$ and $h_{l,\infty}$ with the appropriate loudspeaker gains g_l according to Table 1

$$h_\circ = \sum_{l=1}^{\mathrm{L}} h_{l,\circ}\, g_l \qquad \text{and} \qquad h_\infty = \sum_{l=1}^{\mathrm{L}} h_{l,\infty}\, g_l, \tag{4}$$

$$\mathrm{LF} = \frac{\int\limits_{t_0}^{80\,\mathrm{ms}} h^2_\infty\,\mathrm{d}t}{\int\limits_{0\,\mathrm{ms}}^{80\,\mathrm{ms}} h^2_\circ\,\mathrm{d}t}. \tag{5}$$

As the upper integration bound is normally set to 80 ms, the measure is sometimes also called early lateral energy fraction (Cox *et al.*, 1993). According to ISO3382 (ISO, 2009), the lower integration bound of the figure-of-eight signal is defined as $t_0 = 5$ ms. Although some authors calculate the LF in three octave bands around 500 Hz, 1 kHz, and 2 kHz, i.e. similar to the $\mathrm{IACC_{E3}}$, the broadband version of the LF is used in this article.

Literature gives the following values for the JND of the LF: 0.048 (computed) and 0.058 (measured) (Cox *et al.*, 1993), 0.075 (ISO, 2009), and 0.045–0.07 (Blau, 2002).

The measurements used a Schoeps CCM 8 figure-of-eight microphone and an NTI MM2210 omnidirectional microphone. Figure 5 shows that the standard LF measure is not related ($R^2 = 0.19$) to the listening test results. Furthermore, the LF values range from 0.01 to 0.022 and lie within one JND only, which contradicts the listening test results. These LF values represent the effect of the early reflections exclusively that are perceptually independent of the condition. The exclusion of the direct part of the sound is caused by the lower integration bound of $t_0 = 5$ ms for the figure-of-eight signal.

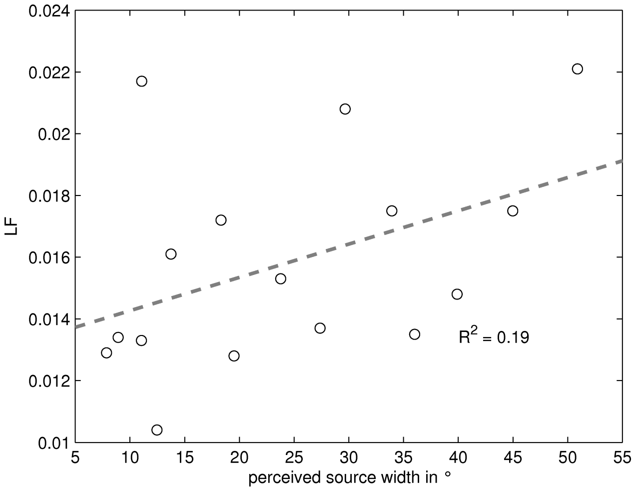

Fig. 5. Regression of the standard LF measures ($t_0 = 5$ ms) to the listening test results.

3.2.1. Reducing the lower integration bound t_0 for h_∞

In order to improve the suitability of the LF for direct sound, the lower integration bound t_0 of the figure-of-eight signal h_∞ is changed to $t_0 = 0$ ms.

Figure 6 shows the regression results of the improved LF measures ($t_0 = 0$ ms) to the listening test results. The value of $R^2 = 0.69$ verifies the improvement of the LF measures. The new measures range from 0.015 to 0.094 and cover 2–3 JNDs. Thus, the quality of the new LF measures is comparable to the $\mathrm{IACC}_{\mathrm{E3}}$ measures.

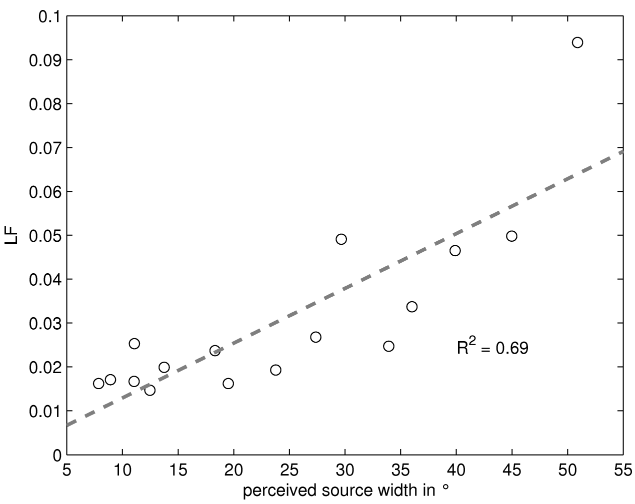

Fig. 6. Regression of the LF measures ($t_0 = 0$ ms) to the listening test results.

The limitation of the range is most likely caused by signal cancellation of the simultaneous sound incidence at the figure-of-eight microphone. The next paragraph presents an approach to overcome the poor prediction of the significant differences from the listening test by further improvement of the LF measurement.

3.2.2. Energetic superposition

If the impulse responses of the loudspeakers have been measured independently, their signals can be superimposed without interference. Formally, this is achieved by energetic superposition

$$h_\circ = \sqrt{\sum_{l=1}^{\mathrm{L}} (h_{l,\circ}\, g_l)^2},$$
$$h_\infty = \sqrt{\sum_{l=1}^{\mathrm{L}} (h_{l,\infty}\, g_l)^2}. \tag{6}$$

This was done for all conditions and the LF measures were calculated again. Figure 9 shows that the correlation between these values and the perceived source width is high ($R^2 = 0.95$). The LF values now range from 0.025 to 0.41 and cover 6–9 JNDs, which is comparable to the listening test results.

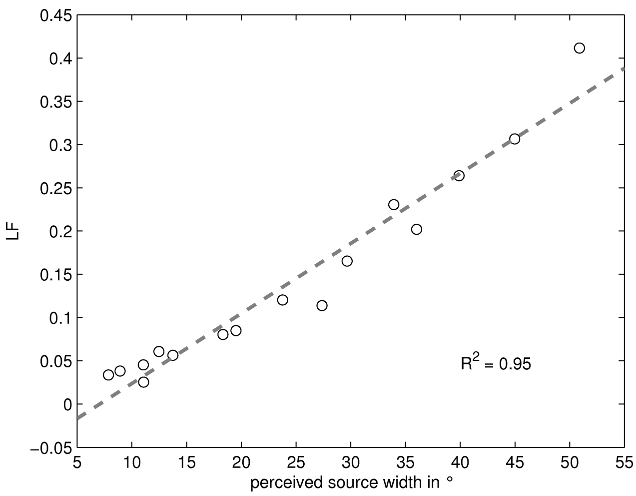

Fig. 7. Regression of the energetically superimposed LF measures ($t_0 = 0$ ms) to the listening test results.

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

3.2.3. Multiple measurement positions

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

$$\mathrm{LF} = \frac{1}{N}\sum_{n=1}^{N}\mathrm{LF}_n. \tag{7}$$

Each of the LF_n values results from a linear superposition of the loudspeaker signals as in Eq. (4). The displacement is only effective above a certain frequency $f_{\min}$ that is determined by the measurement aperture $d_{\max}$, i.e. the distance between the outmost measurement positions

$$f_{\min} = \frac{\mathrm{c}}{2d_{\max}} \tag{8}$$

with c = 343 m/s in air at 20°C temperature.

A simulation was done in order to evaluate the quality of the approximation of the energetic superposition by averaging over multiple measurement positions with linear superposition of the loudspeaker signals. These measurement positions were simulated by appropriate delaying (rounded to integer samples at a sampling rate of 44.1 kHz) and level adjustment of the measured impulse responses from the central listening/measurement position. The simulation ignored the influence of the loudspeaker directivity which is negligible for the used range of $d_{\max}$. In the simulation, the number of measurement positions N and the size of the measurement aperture $d_{\max}$ was varied. The maximum error between the LF measures of the energetic superposition and the average linear superposition at multiple measurement positions was used as quality measure for the approximation. To ensure that there is no perceptible difference in the approximation, the error must be $\leq 1/2$ JND of the LF, which is approximately 0.03. In Fig. 8, the absolute value of the error is presented in gray scale. Whenever the error is below the value of 0.03, the area is white. Darker areas represent larger errors.

The number of measurement positions N has a weak influence on the maximum error. Only for small aperture sizes $d_{\max}$, a larger number N decreases the maximum error. For apertures $d_{\max} \geq 0.14$ m, the errors are smaller than 0.03, even when averaging over only $N = 2$ positions. Interestingly, this distance is

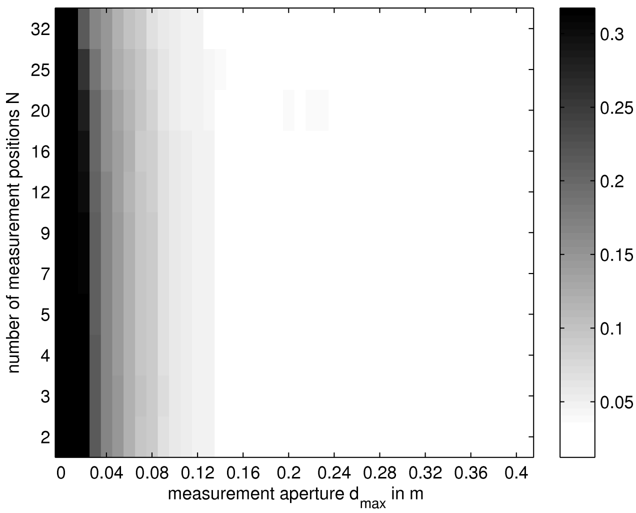

Fig. 8. Maximum error of the approximation of the energetic superposition by averaging over multiple linear superimposed measurement positions in dependence of the number of measurement positions N and the aperture $d_{\max}$.

similar to the head diameter. This distance results in a lower frequency bound for the avoidance of destructive interference of $f_{\min} = 1.2$ kHz. For phantom source width, higher frequency components seem to be important. This result agrees to the findings in (Blauert, Lindemann, 1986b; Morimoto, Maekawa, 1988) stating that higher frequency components also contribute to the perception of source width in concert halls.

Figure 9 shows the regression of the listening test results to the linearly superimposed LF measures that are averaged over $N = 2$ positions with $d_{\max} = 0.14$ m (± 0.07 m apart from the center of the arrangement). The regression yields similar results compared to the energetically superimposed LF measures, cf. Fig. 7.

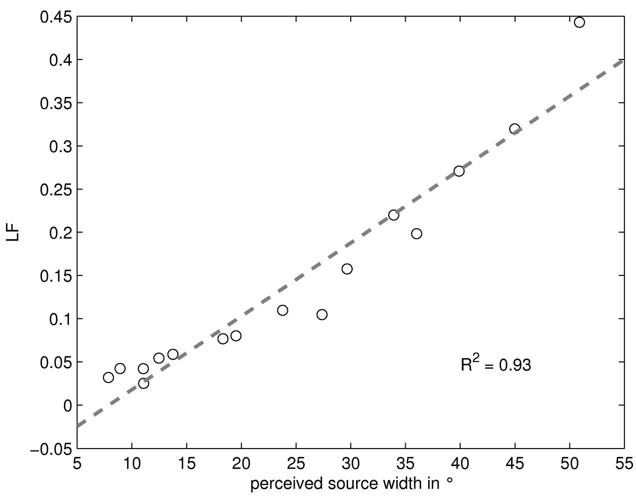

Fig. 9. Regression of the LF measure with linear superposition at 2 positions $d_{\max} = 0.14$ m ($t_0 = 0$ ms) to the listening test results.

Figure 10 compares the different LF measures. The standard LF measure (1 position, linear superposition, $t_0 = 5$ ms) yields small values and does not increase for conditions with larger loudspeaker spacing. Lowering of the integration bound for h_∞ to $t_0 = 0$ ms improves the quality of the LF measure. The largest value range and the best correlation to the listening test results is achieved by the energetic superposition or its approximation by two measurement positions with linear superposition. The advantage of multiple measurement positions can also be found in other fields of acoustics. In recording technique, spaced main microphone arrays are preferred over coincident arrays when capturing spatial impressions (Theile, 1991). The literature about room acoustics tells about large spatial fluctuation of room acoustic measures in concert halls that occur only in measurements, but not in perception (de Vries *et al.*, 2001; van Dorp Schuitman, 2011).

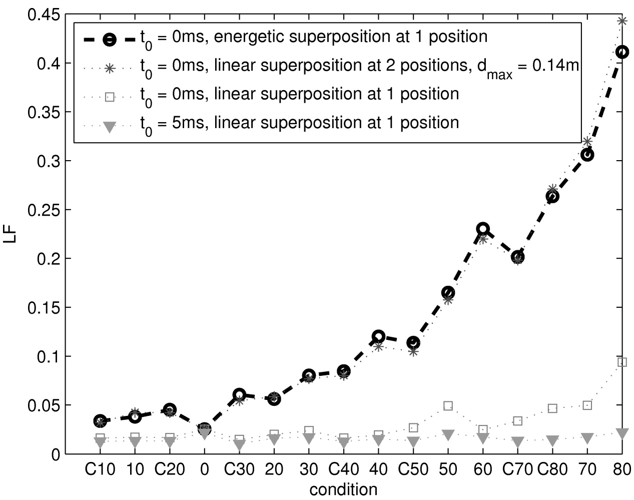

Fig. 10. Comparison of the different LF measures for each condition, arranged to ascending mean values of the listening test results.

Obviously, all presented LF measures yield nearly the same values for condition 0 (a single loudspeaker at 0°). The non-zero values for this condition are due to early reflections and define a lower bound for source width in the listening room. Nevertheless, the differences between the conditions are mainly caused by the differences in the direct sound of the loudspeakers.

The presented adaptations of the LF measure will not decrease their suitability for room acoustic measurements, where source width is mainly caused by early reflections. They will yield similar results as the standard LF measure when they are applied in reverberant rooms, as it is unlikely that there are reflections within the first 5 ms or multiple early reflections arriving at exactly the same time on opposite sides of the figure-of-eight microphone.

4. Simple prediction model

The previous section presented technical measures based on acoustic measurements. These measurements still require the efforts to set up the system under evaluation. This section presents a simple model that can predict the perceived source width without the need of a real setup. The model assumes that the direct sound is more prominent than the early reflections, resulting in a condition-independent contribution of the reflections to the source width. As shown in the section above, this assumption holds for our listening setup, as well as for setups according to the ITU recommendation (ITU, 1997) where the listener sits within the effective critical distance.

4.1. Energy vector ($\boldsymbol{r}_\mathrm{E}$)

The magnitude of the so-called *energy vector* $\boldsymbol{r}_\mathrm{E}$ (Gerzon, 1992) is proposed as predictor of the perceived source width. It is calculated from the direction vectors $\boldsymbol{\theta}_l$ and scalar gains g_l of each loudspeaker

$$\boldsymbol{r}_\mathrm{E} = \frac{\sum\limits_{l=1}^{\mathrm{L}} g_l^2 \boldsymbol{\theta}_l}{\sum\limits_{l=1}^{\mathrm{L}} g_l^2}. \tag{9}$$

In the tested conditions, all L active loudspeakers are driven by $g_l = 1/\sqrt{\mathrm{L}}$ and the overall energy is normalized $\sum\limits_{l=1}^{\mathrm{L}} g_l^2 = 1$. In this case, $\boldsymbol{r}_\mathrm{E}$ is the average of the direction vectors $1/\mathrm{L} \sum\limits_{l=1}^{\mathrm{L}} \boldsymbol{\theta}_l$. Its direction and magnitude relates to the direction and spread of the acoustic energy, respectively. A magnitude value of 1 indicates that only one loudspeaker is active, while one of 0 corresponds to energy distributed in all directions or in opposing directions. The energy vector was originally proposed as a predictor for the direction of phantom sources (Gerzon, 1992). Later works also use it to describe energy distribution (Daniel, 2001). For the tested, frontal conditions, the magnitude of the energy vector is strongly related to the lateral energy fraction under free-field conditions. Therefore, it is not surprising that first hints towards the correlation of the magnitude of the energy vector to the perceived width of phantom sources were discovered (Frank *et al.*, 2011).

Figure 11 shows an excellent correlation ($R^2 = 0.97$) between the listening test results and the magnitude of the energy vector. The regression coefficients yield a formula for the prediction of the perceived source width

$$\alpha = 186.4° \cdot (1 - |\boldsymbol{r}_\mathrm{E}|) + 10.7°. \tag{10}$$

The additive bias of 10.7° relates to the lower bound for the perception of source width that was

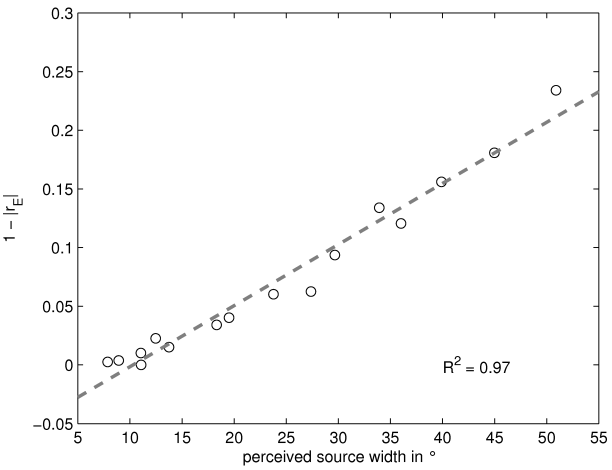

Fig. 11. Regression of $1 - |\boldsymbol{r}_{\mathrm{E}}|$ to the listening test results.

found in the listening test, cf. Fig. 3, and the amount of the LF measure that is caused by early reflections independently of the condition, cf. Fig. 10. Of course, the exact value of this bias is dependent on the room. Nevertheless, the linear relation between the length of the energy vector and the perceived source width is valid for other rooms, as long as the listener sits in the direct sound field of the loudspeakers.

5. Conclusion

This article studied the source width of frontal phantom sources created by two and three loudspeakers playing the same signal in a listening setup with dominant direct sound. For the two-channel conditions, a listening test revealed the relation between the physical width of the active loudspeaker pair and the perceived width of the phantom source. This relation was also demonstrated to work for the case of an additional central loudspeaker. Comparing both groups of stimuli, the addition of the central loudspeaker decreases the source width. This finding agrees with results from the literature (Kin, Plaskota, 2011).

The listening test results were compared to the early inter-aural cross correlation coefficient ($\mathrm{IACC}_{\mathrm{E3}}$) and the lateral energy fraction (LF). The $\mathrm{IACC}_{\mathrm{E3}}$ yields a fair correlation to the listening test results, whereas the correlation to the standard version of the LF is poor. That is because the standard LF measure excludes the direct sound and solely considers the early reflections whose amount was found to be condition-independent in our listening setup. Improvements were presented that adapt the LF measurement for simultaneous incidence of direct sound. By reducing the lower bound for the integration of the figure-of-eight microphone, a fair correlation could be established, which supports the importance of the direct sound. Further improvement could be achieved by avoiding signal cancellation at the figure-of-eight microphone. The optimal solution would be an energetic superposition of the loudspeaker signals at the microphone, which is not possible in practice for simultaneous playback using multiple loudspeakers. The more versatile way of measuring approximates this by measuring at multiple positions. Averaging of two positions at a distance of ± 7 cm is sufficient. As this distance avoids signal cancellation only above 1.2 kHz, the importance of high frequency components for source width is obvious. Despite the improved prediction of source width under listening conditions with dominant direct sound, the adapted LF measures are still applicable for measurements in reverberant rooms.

Finally, a simple model was found that can predict the perceived source width from the listening test solely based on loudspeaker directions and gains. This energy vector model assumes typical studio conditions where the listener sits within the effective critical distance, i.e. the direct sound is more prominent than the reflections. The condition-independent lower bound for source width caused by reflections is incorporated in the model by adding a constant bias. The exact value of this bias depends on the listening setup. The model can be extended to surrounding sound incidence, frequency dependency, and three-dimensional loudspeaker arrangements. Further research is planned about the suitability of the energy vector for predicting phantom source localization.

Acknowledgments

The author is grateful to Franz Zotter for proofreading of the article and fruitful discussions, and the anonymous reviewers for comments that helped improving the paper. This research is partly supported by the project AAP, which is funded by Austrian ministries BMVIT, BMWFJ, the Styrian Business Promotion Agency (SFG), and the departments 3 and 14 of the Styrian Government. The Austrian Research Promotion Agency (FFG) conducted the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a program of the above-mentioned institutions.

References

1. Barron M., Marshall A.H. (1981), *Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure*, Journal of Sound and Vibration, **77**, 2, 211–232.
2. Blau M. (2002), *Difference limens for measures of apparent source width*, Forum Acusticum, Sevilla, Spain.
3. Blauert J., Lindemann W. (1986), *Spatial mapping of intracranial auditory events for various degrees of interaural coherence*, The Journal of the Acoustical Society of America, **79**, 3, 806–813.

4. Blauert J., Lindemann W. (1986), *Auditory spaciousness: Some further psychoacoustic analyses*, The Journal of the Acoustical Society of America, **80**, 2, 533–542.
5. Blauert J. (1983), *Spatial Hearing*, MIT Press.
6. Blumlein A.D. (1958), *British patent specification 394,325 (improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems)*, J. Audio Eng. Soc., **6**, 2, 91–98.
7. Brookes T., Kim C., Mason R. (2007), *An investigation into head movements made when evaluating various attributes of sound*, Preprint 7031, 122nd Conv. Audio Eng. Soc., Vienna.
8. Cox T.J., Davies W.J., Lam Y.W. (1993), *The sensitivity of listeners to early sound field changes in auditoria*, Acta Acustica united with Acustica, **79**, 1, 27–41.
9. Daniel J. (2001), *Représentation de champs acoustiques, application à la transmission et à la reproduction de scénes sonores complexes dans un contexte multimédia*, Ph.D. Thesis, Université Paris 6.
10. de Vries D., Hulsebos E.M., Baan J. (2001), *Spatial fluctuations in measures for spaciousness*, The Journal of the Acoustical Society of America, **110**, 2, 947–954.
11. Frank M., Sontacchi A. (2012), *Performance review of an expert listening panel*, [in:] Fortschritte der Akustik, DAGA, Darmstadt.
12. Frank M., Sontacchi A., Höldrich R. (2010), *Training and guidance tool for listening panels*, [in:] Fortschritte der Akustik, DAGA, Berlin.
13. Frank M., Marentakis G., Sontacchi A. (2011), *A simple technical measure for the perceived source width*, [in:] Fortschritte der Akustik, DAGA, Düsseldorf.
14. Gerzon M.A. (1992), *General metatheory of auditory localisation*, [in:] Preprint 3306, 92nd Conv. Audio Eng. Soc., Vienna.
15. Hidaka T., Beranek L.L., Okano T. (1995), *Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls*, The Journal of the Acoustical Society of America, **98**, 2, 988–1007.
16. ISO (2009), ISO 3382-1:2009: *Acoustics – measurement of room acoustic parameters – part 1: Performance spaces.*
17. ITU (1997), ITU-R BS.1116-1: *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems.*
18. Kin M.J., Plaskota P. (2011), *Comparison of sound attributes of multichannel and mixed-down stereo recordings*, Archives of Acoustics, **36**, 2, 333–345.
19. Mackensen P. (2008), *Auditive Localization & Head Movements*, Ph.D. Thesis, TU Berlin.
20. Martin G., Woszczyk W., Corey J., Quesnel R. (1999), *Controlling phantom image focus in a multichannel reproduction system*, [in:] Preprint 4996, 107th Conv. Audio Eng. Soc., New York.
21. Morimoto M., Iida K. (2005), *Appropriate frequency bandwidth in measuring interaural cross-correlation as a physical measure of auditory source width*, Acoustical Science and Technology, **26**, 2, 179–184.
22. Morimoto M., Maekawa Z. (1988), *Effects of low frequency components on auditory spaciousness*, Acta Acustica united with Acustica, **66**, 4, 190–196.
23. Okano T. (2002), *Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections*, The Journal of the Acoustical Society of America, **111**, 1, 217–229.
24. Pulkki V. (1999), *Uniform spreading of amplitude panned virtual sources*, [in:] IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 187–190.
25. Schroeder M.R., Gottlob D., Siebrasse K.F. (1974), *Comparative study of european concert halls: correlation of subjective preference with geometric and acoustic parameters*, The Journal of the Acoustical Society of America, **56**, 4, 1195–1201.
26. Simon L.S.R., Mason R., Rumsey F. (2009), *Localization curves for a regularly-spaced octagon loudspeaker array*, [in:] Preprint 7915, 127th Conv. Audio Eng. Soc., New York.
27. Sontacchi A., Pomberger H., Höldrich R. (2009), *Recruiting and evaluation process of an expert listening panel*, [in:] NAG/DAGA International Conference on Acoustics, Rotterdam.
28. Theile G. (1991), *On the naturalness of two-channel stereo sound*, J. Audio Eng. Soc., **39**, 10, 761–767.
29. van Dorp Schuitman J. (2011), *Auditory modelling for assessing room acoustics*, Ph.D. Thesis, Technische Universiteit Delft.
30. Wendt K. (1963), *Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie*, Ph.D. Thesis, RWTH Aachen.
31. Zotter F., Frank M., Marentakis G., Sontacchi A. (2011), *Phantom source widening with deterministic frequency dependent time delays*, [in:] DAFx-11, Paris.