Comparison of Interaural Intensity Differences Evoked by Real and Phantom Sources

JEROEN BREEBAART, *AES Member*
(jeroen.breebaart@dolby.com)

Dolby Laboratories, McMahons Point, Australia

This study compares interaural intensity differences (IIDs) of a real source and those resulting from a phantom source created by pair-wise amplitude panning in an anechoic environment with a listener situated in the sweet spot. Factors under investigation are (1) the source frequency, (2) the source direction angle, (3) the loudspeaker angular aperture, (4) the influence of head-related transfer functions (HRTFs) across subjects, and (5) differences between panning laws. The between-subject differences in HRTFs occurred mainly above 1 kHz and were found to be a highly significant factor. For the commonly used loudspeaker angular aperture of 60°, this source of error accounted for 79% of the overall variance. The results also indicated that the most critical direction angle for the evaluation of panning laws equals approximately half that of the loudspeaker angle. Phantom sources tend to have larger IIDs than real sources for the commonly-used loudspeaker angular aperture of 60°, and the magnitude of this offset was found to be frequency dependent. For wider apertures (110°), both larger and smaller phantom-source IIDs were observed, depending on the employed panning law and the frequency of the source signal. Furthermore, substantial errors in IIDs are observed for frequencies at which phase cancellation occurs due to the contribution of two loudspeakers at each ear with a relative time delay. These findings, in relation to the observed between-subject variance, suggest that the accuracy of panning laws can mainly be improved by incorporating frequency and loudspeaker-angle dependent panning functions.

1 INTRODUCTION

When a sound is reproduced by two loudspeakers equidistant from the listener, the resulting phantom source has a perceived direction that is dependent on the (relative) amplitudes of the loudspeaker signals. The effectiveness of such amplitude panning is the result of a translation of inter-channel level differences to interaural time differences (ITDs) and interaural intensity differences (IIDs) at the level of the listener's ear drums, which roughly correspond to those of a real source from the desired sound-source direction [1,2,3,4]. The translation of loudspeaker amplitudes to ITD cues at the level of the ear drums is, however, only accurate at low frequencies [5,6]. At high frequencies (e.g., above 1500 Hz), however, the auditory system is not particularly sensitive to ITDs [7,8] and uses IIDs as dominant azimuth localization cues instead. Similarly to the situation with ITDs, loudspeaker amplitude differences are translated to IIDs as a result of the acoustic shadowing effect of the head, resulting in a fairly accurate match between phantom and real-source IIDs [9].

Amplitude panning has some well-known limitations and drawbacks. For example, the maximum angular aperture (e.g., the angle between the two loudspeakers) for amplitude panning to work accurately in terms of reconstruction of the ITD of a phantom source amounts to approximately 60 degrees [5]. Furthermore, the phantom source created by amplitude panning often has a different perceived timbre compared to a real source from the intended direction, as a result of comb-filter effects [10]. The loudspeaker angular aperture is also quite critical if amplitude panning is employed in front/back directions. If loudspeakers are placed sparsely, the perceived position of the lateral phantom source can be wrong or becomes ambiguous [11,12, 13,14].

Different models (also referred to as panning laws) have been proposed to relate the stereophonic loudspeaker-signal amplitudes to the perceived phantom-source direction, such as the sine law [15,16], the tangent law [17,5], and the sine-cosine law [18]. If the loudspeaker angular aperture is small (60 degrees or less), the sine, sine-cosine, and tangent law differ by only a few degrees in their prediction of the phantom-source direction angle [9]. Data suggest, however, that the perceived phantom-source direction angle depends on the spectrum of the source and is often perceived wider (e.g., having a larger azimuth angle) than predicted by a panning law, especially for frequencies above approximately 1 kHz [2,15, 9,18]. Analyses involving acoustical transfer

functions (head-related transfer functions, or HRTFs) indicated that this discrepancy between predicted and perceived source direction can be explained by considering ITDs and IIDs at the level of the ear drums [9].

It is well known that HRTFs vary considerably across subjects [19,20,21] and that careful selection of HRTFs from a larger set or the use of individualized HRTFs improves the binaural reproduction quality [22,23]. As this variance also influences the sound-source localization cues perceived by the listener, one may hypothesize that for a given set of panning gains and loudspeaker angles, the perceived direction of a phantom source could consequently also be subject dependent. Interestingly, Pulkki and Karjalainen [9] did not find any statistically-significant effect of the subject participating in their test on the judgment of the perceived phantom-source direction. The absence of a statistically significant effect could have had different causes. Perhaps the number of subjects was too small to reject the Null hypothesis as a result of insufficient power or sensitivity. Alternatively, the relatively small loudspeaker angular aperture of 60 degrees may have been insufficient to pull out inter-subject dependencies. Last but not least, even though HRTFs may differ significantly between subjects, the effective localization cues for a phantom and real source may nevertheless be very similar within each subject. In that case, listeners will judge the perceived direction of a real and phantom source similarly despite the differences in localization cues across subjects.

Another research question that seems currently unresolved is which panning law provides the best correspondence between a real source and a phantom source as a function of frequency and loudspeaker angular aperture and how these differences relate to inter-subject variability in HRTF responses. A quantitative analysis of between-subject variance in relation to other factors that influence the performance of panning laws may provide guidance in terms of directions for future panning-law improvements. As far as the author is aware, an explicit comparison of localization cues resulting from phantom and real sources across a set of listeners, different panning laws, source signal frequency, and loudspeaker angular apertures has not been published previously. Given the recent advances in object-based content authoring and reproduction (cf., [24,25]), insight in the performance of panning laws in various conditions may help in improving the spatial imaging capabilities of such systems. This paper therefore compares localization cues resulting from a real source and those from a phantom source. Factors that are being taken into account are a variety of panning laws, the source signal frequency, inter-subject variability in HRTFs, the sound source angle, and the loudspeaker angular aperture. Given earlier reports that differences in the perceived direction of phantom and real sources are typically observed for signals with frequency content above 1 kHz [9,18], we limit this study to interaural intensity differences (IIDs) only. We restrict the analysis to pair-wise, symmetric amplitude panning in the horizontal plane (e.g., without elevation component) with a listener situated in the sweet spot.

This paper is organized as follows. Section 2.1 reviews various panning laws that are used throughout this paper. Section 2.2 describes the method to estimate IID cues of phantom and real sources. The comparison of IID cues for phantom and real sources is outlined in Section 3. A discussion of the results and conclusions are provided in Sections 4 and 5, respectively.

2 METHOD

2.1 Amplitude Panning Laws

Two loudspeaker signals $p_{s,i}(t)$, $i = \{0, 1\}$ are derived from one source signal $p(t)$ as a function of time t using non-negative scale factors referred to as panning gains g_i:

$$p_{s,i}(t) = g_i\, p\,(t)\,. \tag{1}$$

If the azimuth angles of the two speakers are given by $\phi_{s,0} = -\phi_s$; $\phi_{s,1} = +\phi_s$ with $(0 \le \phi_s \le 90^\circ)$, their angular aperture amounts to $2\phi_s$. Throughout this paper, angles and trigonometric functions are expressed in degrees. Under the constraint of equidistant loudspeaker placement with respect to the listener, the predicted phantom source direction angle ϕ_p of the auditory event can be predicted by the stereophonic law of sines derived by [15,16]:

$$\sin(\phi_p) = \sin(\phi_s)\frac{g_1 - g_0}{g_1 + g_0}. \tag{2}$$

Another expression to relate the speaker angle ϕ_s, the scale factors g_i and the phantom-source direction angle ϕ_p is the tangent law [17,5, 9]:

$$\tan\left(\phi_p\right) = \tan\left(\phi_s\right)\frac{g_1 - g_0}{g_1 + g_0}. \tag{3}$$

A popular panning rule in the entertainment industry is the sine-cosine law:

$$g_0 = \cos\left(\alpha\right), \tag{4}$$

$$g_1 = \sin\left(\alpha\right), \tag{5}$$

with α a panning trajectory parameter $(0 \le \alpha \le 90^\circ)$. We then obtain

$$\frac{g_1 - g_0}{g_1 + g_0} = \frac{\sin\left(\alpha\right) - \cos\left(\alpha\right)}{\sin\left(\alpha\right) + \cos\left(\alpha\right)}. \tag{6}$$

Using the sine-cosine summation rule

$$\cos\left(\alpha\right) \pm \sin\left(\alpha\right) = \sqrt{2}\sin\left(45^\circ \pm \alpha\right), \tag{7}$$

we can write

$$\frac{g_1 - g_0}{g_1 + g_0} = \tan\left(\alpha - 45^\circ\right). \tag{8}$$

It is often assumed that the panning parameter α range [0, 90°] is mapped linearly to the phantom source direction angle $\phi_p \subseteq [-\phi_s, \phi_s]$ [18]:

$$\phi_p = \phi_s\left(\frac{\alpha}{45^\circ} - 1\right), \tag{9}$$

resulting in

$$\frac{g_1 - g_0}{g_1 + g_0} = \tan\left(45^\circ\frac{\phi_p}{\phi_s}\right). \tag{10}$$

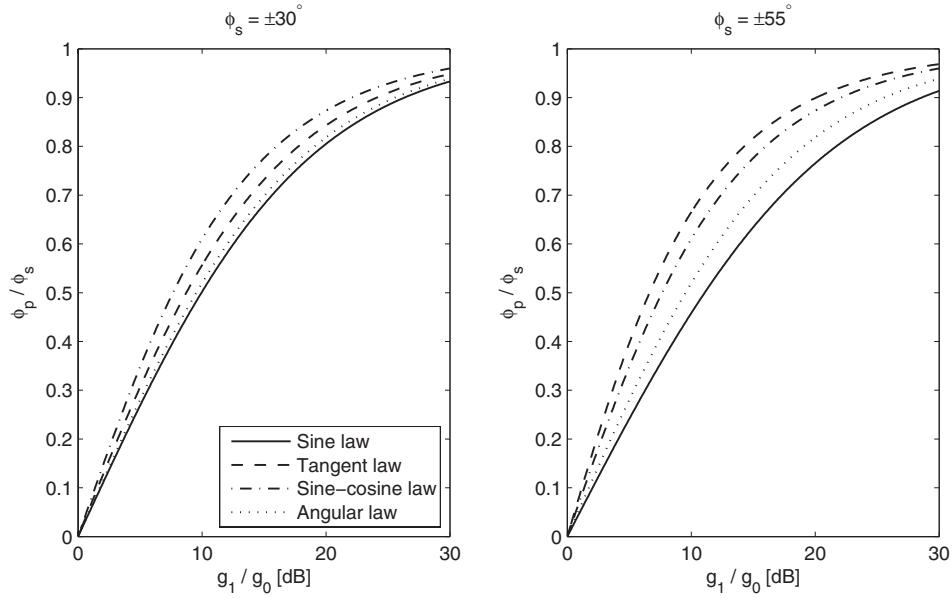

Fig. 1. Normalized predicted panning-law angle ϕ_p/ϕ_s as a function of g_1/g_0 expressed in dB for $\phi_s = 30^\circ$ (left panel) and $\phi_s = 55^\circ$ (right panel). Different line types represent various panning laws (see legend).

Interestingly, for $\phi_s = 45^\circ$ and for the above linear relationship between α and ϕ_p, the sine-cosine law becomes identical to the tangent law. For other loudspeaker angles ϕ_s, this equivalence does not hold.

For small angles ϕ, $\tan(\phi) \approx \sin(\phi) \approx \phi$ and hence one may consider a relatively simple linear relation between amplitude scale factors g_i, the loudspeaker angle ϕ_s, and predicted phantom source direction angle ϕ_p, which we will refer to as an angular panning law:

$$\phi_p = \phi_s \frac{g_1 - g_0}{g_1 + g_0}. \tag{11}$$

Last, but not least, interpolation process can be applied in a Cartesian loudspeaker coordinate system [14]. The unit-distance loudspeaker positions expressed in Cartesian coordinates are given by:

$$x_{s,i} = \sin(\phi_{s,i}), \tag{12}$$

$$y_{s,i} = \cos(\phi_{s,i}). \tag{13}$$

The predicted phantom source position in Cartesian coordinates (x_p, y_p) then results in

$$x_p = \frac{\sum_i x_{s,i} g_i}{\sum_i g_i}, \tag{14}$$

$$y_p = \frac{\sum_i y_{s,i} g_i}{\sum_i g_i}, \tag{15}$$

and consequently

$$\tan(\phi_p) = \frac{x_p}{y_p} = \frac{\sum_i x_{s,i} g_i}{\sum_i y_{s,i} g_i}. \tag{16}$$

For a symmetric loudspeaker setup (e.g., $-\phi_{s,0} = +\phi_{s,1} = \phi_s$) this can also be written as

$$\tan(\phi_p) = \tan(\phi_s)\frac{g_1 - g_0}{g_1 + g_0}, \tag{17}$$

which is the same as the expression for the tangent law given in Eq. 3.

A comparison of the panning laws for $\phi_s = \pm 30^\circ$ and $\phi_s = \pm 55^\circ$ is shown in Fig. 1. The predicted phantom-source direction angle ϕ_p is normalized (e.g., divided by ϕ_s) and is shown as a function of the ratio of the scale factors g_i (expressed in dB). As expected, all panning laws give the same solution for a center direction angle ($\phi_p = 0$; $g_1/g_0 = 0$ dB) and a maximum angle ($\phi_p = \phi_s$; $g_1/g_0 = \infty$). For positions in between, clear differences can be observed, especially for $\phi_s = \pm 55^\circ$. The sine law (solid line) always produces the most conservative phantom-source direction angle. The widest phantom-source direction angle is predicted by the sine-cosine law and tangent law, for $\phi_s = \pm 30^\circ$ and $\phi_s = \pm 55^\circ$, respectively.

2.2 IID Estimation

To simulate the acoustical pathway from the loudspeakers to the eardrums, head-related transfer functions (HRTFs) of the Center for Image Processing and Integrated Computing (CIPIC) database [26] were used. A –3 dB/oct slope was superimposed onto the HRTF spectra as a somewhat crude model for the average slope of audio content, which was shown to be approximately flat when analyzed in third-octave frequency bands between 500 and 5000 Hz [27]. The analysis of IIDs comprises the calculation of ratios of spectrally-averaged energies in auditory filters across ears, and therefore the slope superimposed on the HRTFs spectra determines how fine-structure details in the HRTFs are weighted in the calculation of each averaged auditory-filter signal level and, hence, the resulting IIDs.

For real sources, one HRTF pair was used corresponding to the intended source direction angle. Only azimuth and elevation angle pairs that were present in the HRTF database were used to alleviate the need for HRTF

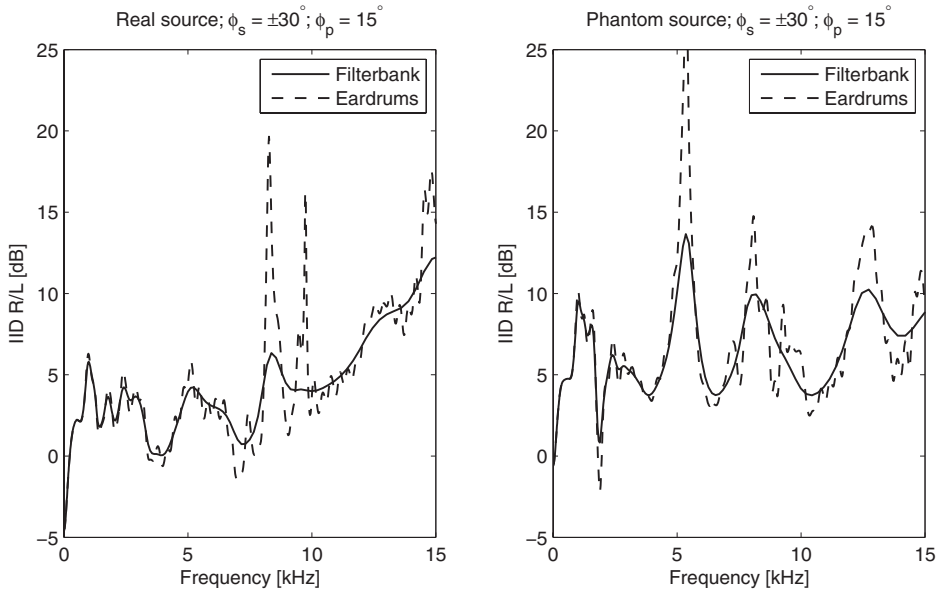

Fig. 2. IID (in dB) as a function of frequency for a real source at $\phi_p = 15^\circ$ (left panel) and a phantom source (right panel) created with loudspeakers at $\phi_s = 30^\circ$ and an angular panning law. The dashed line represents the IID obtained at the level of the ear drums; the solid line is the IID after processing the spectra with a Gammatone filter bank (see text).

interpolation. For phantom sources, two HRTF pairs were used corresponding to the directions of the two (simulated) loudspeakers. The HRTFs were subsequently scaled using the panning gains of the panning law under test and summed to compute the phantom source spectra at the level of the ear drums. An example of the IIDs obtained at the level of the ear drums for CIPIC subject '003', $\phi_s = \pm 30^\circ$ and $\phi_p = 15^\circ$ are visualized by the dashed lines in Fig. 2. The left panel corresponds to a real source, while the right panel shows IIDs for a phantom source. A positive value of the IID indicates a higher level in the right ear. Because the division of the spectra is performed with a high frequency resolution, a narrow notch in the spectrum of the contralateral (left) ear results in a strong peak in the IID curve. Such peaks are observed for both the real as well as the phantom source. It has been argued that for binaural hearing [28] and for HRTFs in particular, fine-structure details in HRTF spectra that can lead to strong resonances in the IID spectrum are, at least for wide-band signals, perceptually irrelevant [29,30] and can be removed by convolving the spectra with an auditory filter bank. A second reason to employ an auditory filter bank at the preprocessing stage is to account for the non-linear frequency resolution of the human auditory system. The estimation of IIDs at the output of an auditory filterbank therefore aims at obtaining a representation that corresponds more closely to the information that the human hearing system uses to localize sound sources [31]. The IIDs used in the analyses in this paper were computed on an Equivalent Rectangular Bandwidth (ERB) resolution by first multiplying the ear-level spectra with the frequency response characteristic of a 4th order Gammatone filterbank with center frequencies spaced at 0.2 ERB and one-ERB bandwidths [32]. Subsequently, the energies in each Gammatone filter of the right and left ears were point-wise divided to compute the interaural intensity difference and were expressed in dB. The resulting IIDs are shown in Fig. 2 by the solid lines.

A direct comparison of IIDs obtained for a real source and a phantom source after processing with a Gammatone filterbank is provided in Fig. 3. The loudspeaker angle ϕ_s was again set to $\pm 30^\circ$, and the sound source direction angle amounted to $\phi_p = 15^\circ$. The loudspeaker panning gains were derived using the angular panning law, and HRTFs were used from subjects '003' and '011' in the CIPIC database, for the left and right panels of Fig. 3, respectively. As can be observed from the figure, the phantom-source IIDs are typically larger (or wider) than those of the real source. Furthermore, the IIDs for subject '003' (left panel) are generally smaller than those for subject '011' (right panel).

2.3 IID Error Statistics

To quantitatively describe the difference in IIDs between phantom and real sources, three different error metrics are evaluated. The first metric is the root-mean square error (RMSE) σ, which is the root of the average squared difference of IIDs between a real and phantom source:

$$\sigma^2(q, \phi_p, \phi_s) = \langle (\lambda(f, j, q, \phi_p, \phi_s) - \lambda'(f, j, q, \phi_p, \phi_s))^2 \rangle, \quad (18)$$

with $\lambda(f, j, q, \phi_p, \phi_s)$ the real source IID expressed in dB and $\lambda'(f, j, q, \phi_p, \phi_s)$ the IID for the corresponding phantom source for frequency f, subject j, panning law q, sound source direction angle ϕ_p, and loudspeaker angle ϕ_s. $\langle . \rangle$ represents the averaging operator applied across all dimensions not included in the arguments of the variable left of the equality sign. The second metric is the mean of the pan-law error μ:

$$\mu(q, \phi_p, \phi_s) = \langle \lambda(f, j, q, \phi_p, \phi_s) - \lambda'(f, j, q, \phi_p, \phi_s) \rangle. \quad (19)$$

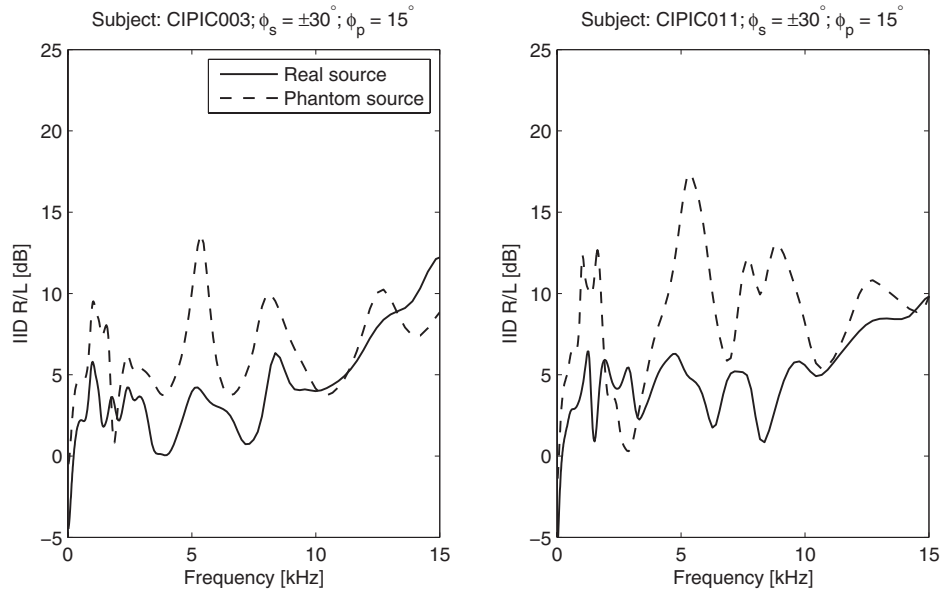

Fig. 3. IID (in dB) as a function of frequency for a real source at $\phi_p = 15°$ (solid line) and a phantom source (dashed line) at the same position (according to the angular panning law) created with loudspeakers at $\phi_s = 30°$. The left and right panels represent different subjects.

Last but not least, the between-subject RMSE ξ is used to quantify variability across subjects and is given by

$$\xi^2(q, \phi_p, \phi_s) = \langle(\lambda(f, j, q, \phi_p, \phi_s) - \lambda'(f, j, q, \phi_p, \phi_s) - \mu(f, q, \phi_p, \phi_s))^2\rangle. \quad (20)$$

To increase the number of experimental observations, and to reduce the effect of potential overall angular offsets that may be present in the HRTF database, mirrored conditions (e.g., with a flipped sign of the source angle ϕ_p and the resulting IIDs) are included in all tests as additional subject indices. With 43 subjects present in the HRTF database, the total number of subject observations therefore amounted to 86.

2.4 Conditions

The following conditions were tested. First, a loudspeaker angle of $\phi_s = \pm 30°$ was employed because this loudspeaker angle is recommended for stereophonic reproduction and for the two front speakers in a 5.1-surround setup [33]. For this condition, the sound source direction angles ϕ_p were spaced linearly from 0 to 25° in 5° steps.

In the second condition, the loudspeaker angle was increased to $\phi_s = \pm 55°$ to investigate the effect of wider loudspeaker angular apertures. In this condition, the sound source direction angles that were tested were linearly spaced from 0 to 45° degrees with a step size of 5°. A sound source angle of ±50° was not included because this sound source position is not included in the CIPIC HRTF database.

3 RESULTS

3.1 Loudspeaker Azimuth Angle of $\phi_s = \pm 30°$

The results for $\phi_s = \pm 30°$ are shown in Fig. 4. The left, middle, and right panels show the overall RMSE σ, the mean panning law error μ and the between-subject RMSE ξ, respectively. In the upper panels, the abscissa reflects the sound source direction angle ϕ_p in degrees and the ordinate the error expressed in dB. The lower panels visualize the error as a function of frequency. Different line styles represent the four panning laws under test (see legend).

The RMSE σ as a function of sound source direction angle ϕ_p (top-left panel in Fig. 4) varies between 2.5 dB (sine law, $\phi_p = 25°$) and 4.6 dB (sine law, ($\phi_p = 10°$), indicating that the most critical direction angles for phantom imaging are 10 and 15°. The differences in the RMSE across panning laws are generally very small with a maximum of about 0.5 dB occurring for $\phi_p = 10°$ in favor of the sine-cosine law. For ϕ_p of 20 and 25°, on the other hand, the sine law has the smallest RMSE. For $\phi_p = 0°$, all panning laws show the same RMSE of approximately 3.2 dB.

The RMSE σ as a function of frequency (bottom-left panel) does not indicate substantial differences between panning laws, with a maximum deviation of 0.6 dB occurring at 5.4 kHz and 0.4 dB above 12 kHz. Furthermore, all panning methods demonstrate a very small RMSE of less than 1.0 dB below 500 Hz, a maximum of about 6.4 dB at approximately 1500 Hz and a local minimum of 2.5 dB at 3.6 kHz. This finding indicates that phantom imaging works very well in the low-frequency range when IIDs are considered, while the most critical frequency region is around 1500 Hz.

The mean pan-law error μ as a function of ϕ_p shown in the top-middle panel of Fig. 4 is zero for $\phi_p = 0°$ and varies

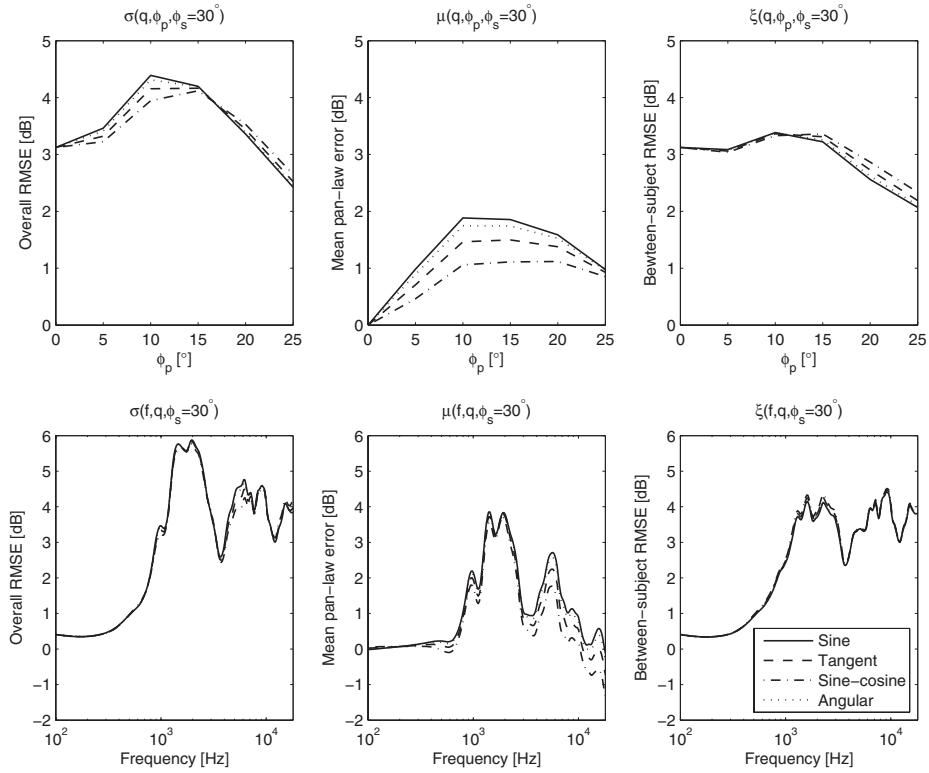

Fig. 4. Overall RMSE (left panel), mean pan-law error (middle panel), and between-subject RMSE (right panel) expressed in dB as a function of source azimuth (upper panels) and frequency (lower panels). Different line types represent different panning laws (see legend). The loudspeaker azimuth angle amounted to $\phi_s = \pm 30^\circ$.

between 1.2 and 2.2 dB for ϕ_p between 10 and 25°. The sine-cosine law has the smallest mean panning-law error, while the sine law shows the largest values.

The lower-middle panel of Fig. 4 indicates that all pan laws show a very small mean panning-law error (less than 0.2 dB) below 600 Hz, an error of up to 3.9 dB between 1.4 and 2.2 kHz and another peak of 2.7 dB at 5.5 kHz. Furthermore, above 3 kHz, the panning laws have very similar shapes except for a vertical shift of up to 1.0 dB. All panning laws under test demonstrate a phantom-source IID that is (on average) larger than a real-source IID, in line with findings reported by [9]. This discrepancy is generally smallest for the sine-cosine panning law, and hence the overall RMSE is smallest for that panning law, except for ϕ_p equal to 20° and 25°.

The between-subject RMSE ξ given in the right panels of Fig. 4 fluctuates between 2.1 and 3.5 dB depending on the sound source direction angle ϕ_p and the panning law under test. The lowest between-subject RMSE is observed for the sine law at $\phi_p = 25^\circ$. The lower panel indicates that ξ is very small (less than 1.0 dB) for frequencies below 500 Hz and varies between 2.3 and 4.3 dB above 1 kHz.

A four-way analysis of variance was carried out on the difference between phantom and real source IIDs, with subject, frequency, sound source direction angle, and panning law as independent factors, including their first-order interactions. All main effects and their interactions were found to be statistically significant ($p < 10^{-6}$) except for the interaction of subject and panning law ($p = 1.0$) and the interaction of frequency and panning law ($p > 0.33$).

3.2 Loudspeaker Azimuth Angle of $\phi_s = \pm 55°$

The results for $\phi_s = \pm 55^\circ$ are shown in Fig. 5. The format of the figure is the same as that of Fig. 4. The RMSE as a function of source direction angle ϕ_p varies between 2.8 and 4.9 dB. Depending on the employed panning law, the maximum occurs between 20 and 30°. The RMSE frequency dependency, as visualized in the lower-left panel of Fig. 5 indicates very small RMSE errors (<0.6 dB) below 300 Hz, and local maxima around 860 Hz, 3200 Hz, and 6800 Hz depending on what panning law is used. Furthermore, up to 1800 Hz, the tangent law has the smallest RMSE while the sine law has the largest RMSE. For higher frequencies, this relationship is reversed with the exception of the local maximum around 3200 Hz.

The mean pan-law error is visualized in the middle panels of Fig. 5. The error is zero for $\phi_p = 0$, as indicated in the top-middle panel. The differences between panning laws are largest for $\phi_p = 30^\circ$. For this angle, the sine and tangent panning law are associated with the highest and lowest pan-law error, respectively. The lower-middle panel indicates that differences between panning laws predominantly exist above 1 kHz.

The two right panels showing the between-subject variance as a function of ϕ_p (upper panel) and frequency (lower

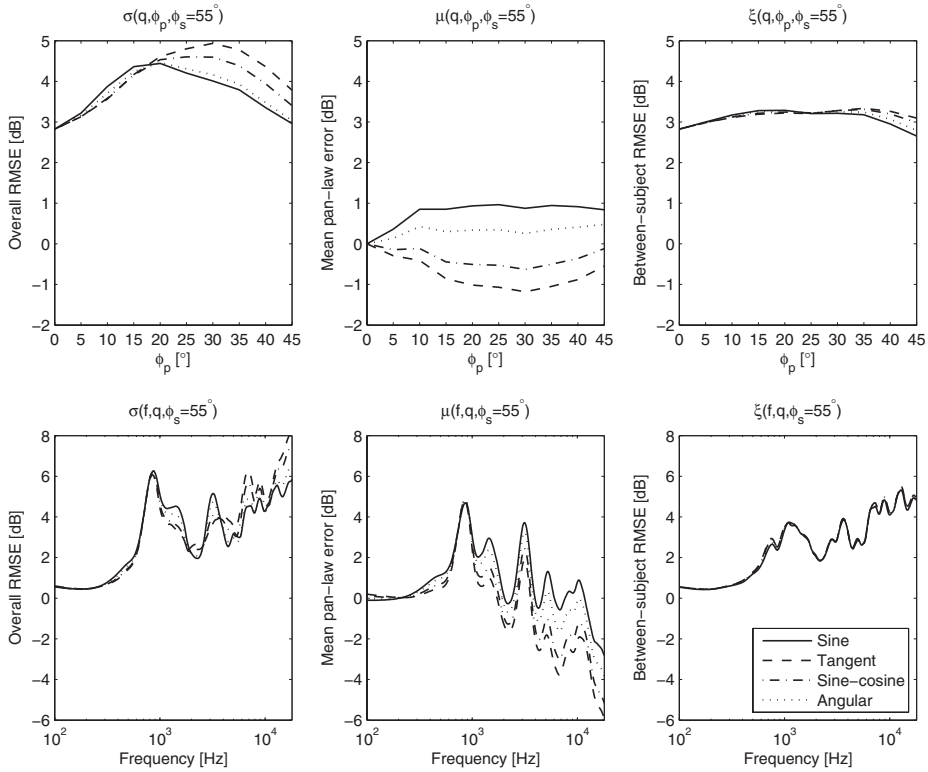

Fig. 5. Overall RMSE (left panel), mean pan-law error (middle panel), and between-subject RMSE (right panel) expressed in dB as a function of source azimuth (upper panels) and frequency (lower panels). Different line types represent different panning laws (see legend). The loudspeaker azimuth angle amounted to $\phi_s = \pm 55^\circ$.

panel) do not indicate any substantial dependencies on the source direction angle ϕ_p nor the panning law. Furthermore, the between-subject variance increases with frequency; it is very small (<0.6 dB) below 300 Hz and fluctuates between 2 and 5 dB for higher frequencies.

A four-way analysis of variance was carried out on the difference between phantom and real source IIDs, with subject, frequency, source direction angle, and panning law as independent factors, including their first-order interactions. All main effects and their interactions were found to be statistically significant ($p < 10^{-6}$) except for the interaction of subject and panning law ($p = 1.0$).

4 DISCUSSION

For $\phi_p = 0^\circ$ (e.g., the source is directly in front of the listener), the employed symmetry that was enforced in the HRTF set results in a mean panning-law error μ of zero, and hence the overall RMSE is equal to the between-subject RMSE. When the source angle is increased to approximately half the loudspeaker angle (e.g., $\phi_p = \phi_s/2$), the overall RMSE increases to its maximum as a result of mean panning-law error contributions. This finding suggests that this angle is the most critical angle to evaluate (differences in) panning laws. A further increase of ϕ_p toward ϕ_s results in a decrease of all three error metrics. Such decreases are expected because in the limit case $\phi_p = \phi_s$, the phantom and real source result in identical signals arriving in the eardrums and hence all error metrics converge to a value of zero. Interestingly, for $\phi_s = \pm 55^\circ$, the source angle region between $\phi_p = 25^\circ$ and 45° shows considerable differences in the overall RMSE between panning laws. This is likely the result of the fact that the panning laws differ significantly for those source direction angles (cf., the right panel in Fig. 1).

Low frequencies do not evoke substantial IIDs due to the relatively large wavelength compared to the size of the head and consequently an absence of an acoustic shadowing effect. Furthermore, the contributions of both loudspeakers to each ear in a phantom-source condition will sum up virtually coherently due to the small differences in arrival times. Consequently, the overall RMSE, the mean panning-law error and the between-subject RMSE are all very small below 500 Hz. If one employs a somewhat arbitrary threshold of 1.0 dB upon the mean panning-law error, which is about twice the magnitude of a just-noticeable difference in IID [34,35], the upper frequency for which small errors are obtained amount to approximately 800 and 500 Hz, for $\phi_s = 30^\circ$ and 55°, respectively.

The largest difference between phantom and real source IID is observed for frequencies corresponding to half the arrival time differences of the two loudspeakers at the level of the ear drum, resulting in phase cancellation. For a $\phi_s = \pm 30^\circ$ and a spherical head model with a head radius of 9.5 centimeters [36], the ITD amounts to about 289 μs [37] and therefore the lowest two frequencies at which phase cancellation is expected to occur are equal to 1.73 kHz and 5.19 kHz. The same phenomenon occurs for

$\phi_s = \pm 55^\circ$, albeit at lower frequencies (0.97 and 2.91 kHz) due to the larger ITD (516 μs) associated with the larger loudspeaker angle. The net effect of this cancellation for off-center source directions is that the IID of the phantom source is larger than the IID of the corresponding real source. This is reflected as a peak in both the mean panning law error and the overall RMSE.

For frequencies above the first phase-cancellation RMSE peak (and ignoring the second peak), the mean panning-law error gradually decreases with frequency. This is because the head acts as an acoustic shadow, effectively translating inter-channel level differences in a phantom-source condition into IIDs at the level of the ear drum. As the magnitude of the acoustic shadow effect increases with frequency, the resulting IIDs for both a real and a phantom source increase with frequency as well, and this effect seems to occur faster with frequency for the real source than for the phantom source. These findings support earlier statements by others [15,9,18] that the accuracy of panning laws depends on the frequency of the source. More specifically, these results confirm earlier statements (cf., [15]) that panning laws can benefit from a frequency-dependent panning gain ratio to compensate for the above effect and to produce a mean panning-law error that is constant (and preferably small) with frequency.

The between-subject RMSE was found to be small at low frequencies, because neither phantom nor real sources evoke substantial IIDs irrespective of the source direction angle. For frequencies above 1 kHz, on the other hand, the between-subject RMSE contributes significantly to the overall RMSE. Its variation amounts to approximately 2 to 5 dB across a wide range of source direction angles. In fact, for $\phi_s = \pm 30^\circ$, the between-subject RMSE accounts for 79% of the overall variance in IIDs, while it accounts for approximately 64% of the variance for $\phi_s = \pm 55^\circ$. This quantification of between-subject variance may change the perspective on what aspects of a panning law are considered to be important and what directions for improvement would be worthwhile to pursue. In particular, the results may provide a concurrent viewpoint on earlier statements that differences between panning laws are considered to be important [18]. The current data indicate that, at least for $\phi_s = \pm 30^\circ$, the differences in IIDs evoked by various panning laws are very small compared to the between-subject variance and, in fact, are often smaller than the just-noticeable differences in IIDs of approximately 0.5 dB [34,35]. On the other hand, the sound source frequency and the loudspeaker angle have shown to be considerable factors in determining the phantom-source IID in relation to the real source IID, suggesting that panning laws should preferably operate in a frequency and loudspeaker-angle dependent manner.

4.1 Limitations

The results reported in this paper are subject to some limitations. First, the analysis only addresses IID localization cues in critical bands. Although there is substantial evidence that these are the most relevant cues for sound source localization above 1.5 kHz, the combination of information present in multiple frequency bands into one source localization precept is essentially missing. For broad-band signals, such a spectral integration process could effectively remove part of the IID variance observed in this study, and, therefore, some of the observations and conclusions in this study may change when the perceived direction would have been used as response variable. In relation to this, ITD cues were not included in this analysis, which may substantially influence the perceived location of a source, especially if the source signal contains a substantial amount of energy below 1.5 kHz. This means that the results reported here are mostly valid for high-pass stimuli in which localization is dominated by IIDs.

A second limitation relates to the fact that only anechoic conditions were simulated. The presence of room reflections and late reverberation is likely to have an effect on the IID estimates. Moreover, the resulting change in interaural correlation may change the way localization cues are integrated across frequency such as proposed by [38].

A third limitation is that IID cues were only analyzed assuming a listener situated in the sweet spot. In practice, this is almost never the case. Additionally, head translations and rotations or slightly misplaced loudspeakers may have further effects on IID cues resulting in deviations in IID cues that could potentially be larger than the effects shown here.

5 CONCLUSIONS

The frequency-dependent acoustic shadow effect of the human head gives rise to frequency-dependent IIDs for both real as well as phantom sources. An ideal panning law would preferably evoke identical IIDs in both conditions. The results of this study indicate that the translation of panning gain ratios to IIDs depends on the source frequency, the individual's HRTFs, the loudspeaker angle, and the source direction angle. For small loudspeaker angular apertures (e.g., 60°), the IIDs of a phantom source are typically larger in absolute sense than those of a real source from a direction predicted by the panning laws under test, especially above 1 kHz. Consequently, the most conservative panning law (e.g., the sine-cosine law producing the smallest panning gain ratio for a given source direction) generally results in the best correspondence of phantom and real source IIDs. For wider loudspeaker angular apertures (110°), however, the IID of the phantom source is either smaller or larger than the IID of the real source, depending on the sound source frequency and the panning law under test. When considering the effect of the sound source direction angle, the maximum deviation in IIDs occurs for a sound source direction angle in the vicinity of half the loudspeaker angle, suggesting that this source angle is most critical for the evaluation of amplitude panning laws. Last, but not least, the translation of panning gains to IIDs is subject to differences in HRTFs among listeners in all conditions. This subject dependency occurs mainly above 1 kHz and showed to account for 79 and 64% of the overall variance, for a loudspeaker angular aperture of 60° and 110°, respectively. Given the considerable magnitude

of this effect in relation to the relatively small differences observed when different wide-band panning laws are compared, the results suggest that panning laws can mostly be improved by allowing frequency and loudspeaker-angle dependent operation.

6 ACKNOWLEDGMENT

The author would like thank the reviewers for their helpful comments to improve the manuscript.

REFERENCES

[1] A. D. Blumlein, Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems, British patent no. 34657 reprinted in *Stereophonic Techniques* (Audio Engineering Society, New York, 1986), 1933.

[2] K. de Boer, *Stereofonische geluidsweergave*, PhD thesis, Technische Hoogeschool, Delft, The Netherlands, April 1941.

[3] S. P. Lipshitz "Stereo Microphone Techniques: Are the Purists Wrong?"*J. Audio Eng. Soc.*, vol. 34, pp. 716–744 (1986 Sep.).

[4] E. Sengpiel, Höreignisrichting b1 in Abhängigkeit von der Interchannel-Pegeldifferenz ΔL. http://www.sengpielaudio.com/HoerereignRichtungDL.pdf, March 1994, retrieved August 6, 2013.

[5] J. C. Bennett, K. Barker, and F. O. Edeko, "A New Approach to the Assessment of Stereophonic Sound System Performance," *J. Audio Eng. Soc.*, vol. 33, pp. 314–321(1985 May).

[6] E. Benjamin and P. Brown, "The Effect of Head Diffraction on Stereo Localization in the Mid-Frequency Range," presented at the *122nd Convention* of the Audio Engineering Society (2007 May), convention paper 7018.

[7] R. G. Klumpp and H. R. Eady, "Some Measurements of Interaural Time Difference Thresholds,"*J. Acoust. Soc. Am.*, vol. 28, pp. 859–860 (1956).

[8] F. L. Wightman and D. J. Kistler, "The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization," *J. Acoust. Soc. Am.*, vol. 91, pp. 1648–1661 (1992).

[9] V. Pulkki and M. Karjalainen, "Localization of Amplitude-Panned Virtual Sources. I: Stereophonic Panning," *J. Audio Eng. Soc.*, vol. 49, pp.739–752 (2001 Sep.).

[10] V. Pulkki, "Coloration of Amplitude-Panned Virtual Sources," presented at the *110th Convention* of the Audio Engineering Society (2001 May), convention paper 5402.

[11] G. Theile and G. Plenge, "Localization of Lateral Phantom Sources,"*J. Audio Eng. Soc.*, vol. 25, pp. 196–200 (1977 Apr.).

[12] G. Martin, W. Woszczyk, J. Corey, and R. Quesnel, "Sound Source Localization in a Five-Channel Surround Sound Reproduction System," presented at the *107th Convention* of the Audio Engineering Society (1999 Sep.), convention paper 4994.

[13] V. Pulkki, M. Karjalainen, and V. Valimaki, "Localization, Coloration and Enhancement of Amplitude-Panned Virtual Sources," *Proceedings 16th AES Conference* on Spatial Sound Reproduction (1999 Mar.), paper 16-024, pages 257–278.

[14] V. Pulkki and M. Karjalainen, "Localization of Amplitude-Panned Virtual Sources. II: Two- and Three-Dimensional Panning," *J. Acoust. Soc. Am.*, vol. 49, pp. 753–767 (2001 Sep.).

[15] H. A. M. Clark, G. F. Dutton, and P. B. Vanderlyn, "The 'Stereosonic' Recording and Reproducing System," *J. Audio Eng. Soc.*, vol. 6, pp. 102–117 (1958 Apr.).

[16] B. B. Bauer, "Phasor Analysis of Some Stereophonic Phenomena," *J. Acoust. Soc. Am.*, vol. 33, pp. 1536–1539 (1961 Nov.).

[17] D. M. Leakey, "Some Measurements on the Effect of Interchannel Intensity and Time Differences in Two Channel Sound Systems," *J. Acoust. Soc. Am.*, vol. 31, pp. 977–986 (1959 July).

[18] D. Griesinger, "Stereo and Surround Panning in Practice," presented at the *112th Convention* of the Audio Engineering Society (2002 Apr.), convention paper 5564.

[19] D. J. Kistler and F. L. Wightman, "A Model of Head-Related Transfer Functions Based on Principal Components Analysis and Minimum-Phase Reconstruction," *J. Acoust. Soc. Am.*, vol. 91, pp. 1637–1647 (1992).

[20] F. L. Wightman and D. J. Kistler, "Individual Differences in Human Sound Localization Behavior," *J. Acoust. Soc. Am.*, vol. 99, pp. 2470–2500 (1996).

[21] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?"*J. Audio Eng. Soc.*, vol. 44, pp. 451–469 (2006 June).

[22] F. Rumsey, "Whose Head Is it Anyway? Optimizing Binaural Audio," *J. Audio Eng. Soc.*, vol. 59, pp. 672–675 (2011 Sep.).

[23] D. Schönstein and B. F. G. Katz, "Variability in Perceptual Evaluation of HRTFs," *J. Audio Eng. Soc.*, vol. 60, pp. 783–793 (2012 Oct.).

[24] Dolby Laboratories, Dolby Atmos—Next Generation Audio for Cinema. http://www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/Dolby-Atmos-Next-Generation-Audio-for-Cinema.pdf, November 2012. Retrieved August 6, 2013.

[25] J. Herre, H. Purnhagen, J. Koppens, O. Hellmuth, J. Engdegård, J. Hilpert, L. Villemoes, L. Terentiv, C. Falch, A. Hölzer, M. L. Valero, B. Resch, H. Mundt, and H.-O. Oh, "MPEG Spatial Audio Object Coding—The ISO/MPEG Standard for Efficient Coding of Interactive Audio Scenes," *J. Audio Eng. Soc.*, vol. 60, pp. 655–673 (2012 Sep.).

[26] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pp. 99–102, New Paltz, NY, Oct 21–24 2001. Mohonk Mountain House.

[27] P. J. Chapman, "Programme Material Analysis," presented at the *100th Convention* of the Audio Engineering Society (1996 May), convention paper 4277.

[28] H. S. Colburn and N. I. Durlach, "Models of Binaural Interaction," in E. Varterette and M. Friedman, editors,

Handbook of Perception, volume IV, pp. 467–518 (Academic, New York, 1978).

[29] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Sensitivity of Human Subjects to Head-Related Transfer-Function Phase Spectra," *J. Acoust. Soc. Am.*, vol. 105, pp. 2821–2840 (1999 May).

[30] J. Breebaart, F. Nater, and A. Kohlrausch, "Spectral and Spatial Parameter Resolution Requirements for Parametric, Filter-Bank-Based HRTF Processing," *J. Audio Eng. Soc.*, vol. 58, pp. 126–140 (2010 Mar.).

[31] J. Blauert, *Spatial Hearing—The Psychophysics of Human Sound Localization*, volume 2nd enhanced edition (The MIT Press, USA-Cambridge MA, 1997).

[32] B. R. Glasberg and B. C. J. Moore, "Derivation of Auditory Filter Shapes from Notched-Noise Data," *Hearing Research*, vol. 47, pp. 103–138 (1990).

[33] ITU, *Multichannel Stereophonic Sound System With and Without Accompanying Picture (ITU-R Recommendation BS.775-2)* (International Telecommunications Union, July 2006).

[34] A. W. Mills, "Lateralization of High-Frequency Tones," *J. Acoust. Soc. Am.*, vol. 32, pp. 132–134 (1960).

[35] R. C. Rowland Jr. and J. V. Tobias, "Interaural Intensity Difference Limen," *J. Speech and Hearing Research*, vol. 10, pp. 733–744 (1967).

[36] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a Spherical-Head Model from Anthropometry," *J. Audio Eng. Soc*, vol. 49, pp.472–479 (2001 June).

[37] R. S. Woodworth and G. Schlosberg, *Experimental Psychology* (Holt, Rinehard and Winston, New York, 1962), pp. 349–361.

[38] C. Faller and J. Merimaa, "Source Localization in Complex Listening Situations: Selection of Binaural Cues Based on Interaural Coherence," *J. Acoust. Soc. Am.*, vol. 116, pp. 3075–3089 (2004).

THE AUTHOR

Jeroen Breebaart

Jeroen received an MSc degree in biomedical engineering from the Eindhoven University of Technology in 1997, and a PhD degree in auditory psychophysics from the same university in 2001. From 2001 to 2007, Jeroen was with the Digital Signal Processing group at Philips Research, conducting research in the areas of spatial perception, stereo and multi-channel parametric audio coding, automatic audio content analysis and binaural rendering. From 2007 to 2010 Jeroen was with the Information and System security group at Philips Research, and worked in the areas of biometric information protection and automated human behavior analysis. He actively participated in the ISO/IEC IT security techniques standardization committee (JTC1 SC27) as co-editor of project 24745 (Biometric information protection). In 2011 and 2012 Jeroen was with Civolution, developing watermarking algrithms for broadcast monitoring, second screen applications and forensic tracking of media content. In 2012, Jeroen moved to Sydney, Australia to join Dolby Laboratories.