COMPENSATING DISPLACEMENT OF AMPLITUDE-PANNED VIRTUAL SOURCES

VILLE PULKKI

Laboratory of Acoustics and Audio Signal Processing, P.O.Box 3000, FIN-02015 HUT, Finland
Ville.Pulkki@hut.fi

The localization of amplitude-panned virtual sources is biased towards the median plane when loudspeakers are not positioned symmetrically with the median plane. The bias is measured using listening tests and with computational modeling of virtual source perception. A modification to an existing pair-wise panning method is proposed that compensates the displacement of virtual sources. The proposed method is evaluated by interpreting conducted listening test results, and by simulating virtual sources with computational modeling. Evaluations suggest that the bias is non-existing with the proposed method.

INTRODUCTION

Multi-channel loudspeaker systems are used to create immersive auditory displays [1, 2]. Amplitude panning is a method to position virtual sources in such displays. A virtual source is a perception of sound source that does not coincide with any physical sound source. In earlier studies it has been shown that the positions of virtual sources are controlled with good accuracy when the loudspeakers are placed symmetrically with the median plane of a listener. In such cases the panning direction corresponds well to the perceived direction [3]. In other cases virtual source direction is biased towards the median plane of the listener, when compared with panning direction [4].

A panning method that would compensate the bias would be beneficial. Compensation is possible in applications in which the orientation of a listener is known. This holds in various cases, e.g. computer users and movie watchers seldom look elsewhere than at the screen. However, sometimes compensation cannot be performed. In immersive visual displays the orientation of a user is not limited. In such cases compensation is possible only if the listener's head position and orientation is tracked.

This paper measures the virtual source direction bias using subjective listening tests and computational modeling of directional hearing mechanisms. Listening tests are conducted by using synthetic and natural broad-band sounds to explore the need of compensation with typical listening material. Using this data, an enhancement to the vector base amplitude panning (VBAP) [5] method is suggested, that compensates the bias at lateral directions.

This paper is organized as follows. The directional hearing mechanisms of humans are reviewed in section 1 and amplitude panning techniques in Section 2. Conducted listening tests are described in Section 3, and a method to enhance VBAP is presented in Section 4.

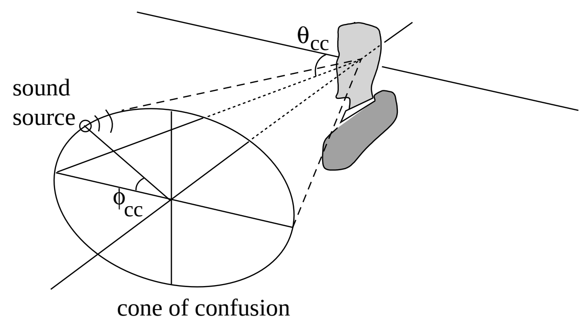

Figure 1: Cone of confusion. Spherical coordinate system that is convenient in directional hearing. The direction of a sound source is denoted as (θ_{cc},ϕ_{cc}) pair.

1. DIRECTIONAL HEARING

Humans are able to perceive directions of sound sources, or more generally, directions of auditory objects. An auditory object is an internal percept that is caused by a sound source emitting sound signals (or sound waves). In this Section the main concepts and mechanisms in directional hearing are reviewed.

Before perceptual issues are considered, some definitions are in order. The median plane divides symmetrically the space related to a listener into left and right parts. The plane that divides space into upper and lower parts is called the horizontal plane. A cone of confusion [6] is defined as a set of points which all satisfy following condition, the difference of distances from both ears to any point on the cone is constant. It can be approximated by a cone having axis of symmetry along a line passing through the listener's ears and having the apex in center point between the listener's ears, as in Fig. 1.

Spherical coordinates azimuth (θ) and elevation (ϕ) are often used to denote sound source directions in research of spatial hearing. This is, however, not a convenient coordinate system, since a cone of confusion can not be specified easily. An alternative spherical coordinate system has been used by Duda [7]. The sound source location is specified by defining the cone in which it lies, and further by the direction within the cone. The cone is defined by the angle θ_{cc} between the median plane and the cone, Fig. 1. Variable θ_{cc} may have values between $-90°$ and $90°$. The direction within the cone of confusion is denoted as ϕ_{cc}, and it varies between $-180°$ and $180°$.

1.1. Binaural cues

Temporal differences between ear canal signals are called the interaural time differences (ITD) and magnitude differences are called the interaural level differences (ILD) [6, 8]. These differences are caused respectively by the wave propagation time difference (primarily below 1.5 kHz) and the shadowing effect by the head (primarily above 1.5 kHz). When a sound source is shifted within a cone of confusion, ITD and ILD remain constant, although, there might be some frequency-dependent changes in ILD [7]. In (θ_{cc},ϕ_{cc}) spherical coordinate system this means that θ_{cc} can be decoded with ILD or ITD. The auditory system decodes the cues in a frequency-dependent manner.

In psychoacoustic tests it has been found that mechanisms that decode ITD are sensitive to the phase of signal at low frequencies below roughly 1.6 kHz and to envelope time shifts at high frequencies [6]. ITD is a quite stable cue; its value is almost constant with frequency. The absolute value, however, is higher at low frequencies. Also, the dependence of ITD on θ_{cc} is monotonic.

It has been found that ILD mechanisms are sensitive to differences of sound pressures at ear canals. Due to complex wave acoustics around the head, ILD is largely dependent on frequency: it is negligible at low frequencies and increases nonmonotonically with frequency. The behaviour of ILD with θ_{cc} is also problematic. It behaves monotonically only within some region $-\gamma < \theta_{cc} < \gamma$ [6] depending on frequency, where the value of γ is typically $40° - 80°$. ILD is also dependent on distance of a sound source [9]. However, in this study sound sources are in far field, where this effect can be neglected.

1.2. Additional cues

Binaural cues resolve the θ_{cc} direction of a sound object, or the cone of confusion of it. Additional cues are needed to also resolve the ϕ_{cc} direction within that cone. The pinna causes strong direction-dependent spectral effects to ear canal signals, which are used to decode the ϕ_{cc} direction [6]. Also, the effect of head rotation to binaural cues is used in ϕ_{cc} decoding [6].

The precedence effect [6, 10, 11] is an assisting mechanism of spatial hearing. It is a suppression of early delayed versions of the direct sound in source direction perception. This helps in reverberant rooms to localize sound sources.

1.3. Modeling of virtual source directional perception

In this study amplitude-panned virtual sources are analyzed. In the analysis the precedence effect is not modeled. In principle, the results are thus valid only in anechoic listening conditions. However, since the precedence effect suppresses the effect of room reflections in localization, a listener resolves direction mostly using direct sound. Direct sound is equal in different listening conditions. Thus the analysis is valid also at least for moderately reverberant listening conditions.

In the auditory model used in this study the ear canal signals were simulated with measured head-related trasfer functions (HRTFs), 6 individual HRTF sets were used. The middle ear was modeled with a filter, that approximates a response function derived from the minimum audible pressure curve [12]. The frequency resolution of the cochlea was simulated with a gammatone filter bank with 42 frequency bands [13]. Hair cells were modeled by half-wave rectification and low-pass filtering.

The ITD between ear signals were simulated with a model that is based on the coincidence-counting model by Jeffress [14]. Cross-correlations with different time lags between ear signals were calculated. The time lag that produces the highest value for cross-correlation was considered as the ITD value. ITD was calculated at different frequency bands, which produced values as a function of frequency. It has also been shown that if ITD has the same value at adjacent frequency bands, it is considered more relevant in localization [15]. To simulate this, a second-level coincidence counting unit was added [15]. The cross-correlation function at a frequency band was multiplied with cross-correlation functions at the adjacent bands.

The ILD was modeled by calculating the loudness level of each frequency band and by subtracting the values from the corresponding value at the contralateral ear [16]. The difference of loudness levels between the ears at each frequency band was treated as the ILD spectrum.

ITD cues are measured in milliseconds and ILD cues in phons. This makes it impossible to compare them together. Additionally, individuals may have different ITD and ILD values for a sound source in a same direction. To make it possible to compare the cues together and between individuals, the cue values at each frequency band were translated with a database search to θ_{cc} angles that they suggested, and the final values were called the ITD angle (ITDA) and the ILD angle (ILDA). The database

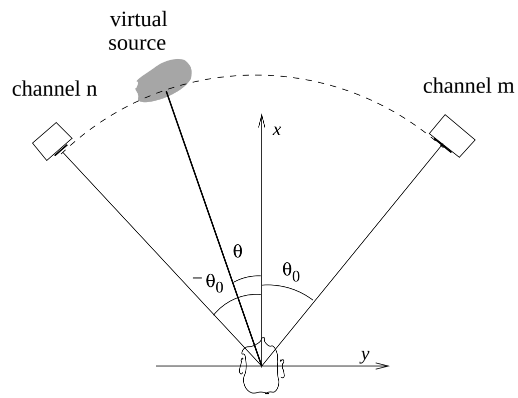

Figure 2: Standard stereophonic listening configuration.

was formed from a large set of cues of real sources in known directions. Details of the models are presented more thoroughly in [3].

The simulations in this study are run by using pink noise as input to the model. Pink noise is considered to excite all frequency bands and both ITD and ILD mechanisms evenly. Due to simplicity no perceptual weighting is performed to the cues, rather medians and quartiles over frequency is performed.

2. AMPLITUDE PANNING

Amplitude panning is the most frequently used technique to position virtual sources. In it a sound signal is applied to loudspeakers with different amplitudes, which can be formulated as

$$x_i(t) = g_i x(t), \quad i = 1, \ldots, N, \tag{1}$$

where $x_i(t)$ is the signal to be applied to loudspeaker i, g_i is the gain factor of the corresponding channel, N is the number of loudspeakers, and t is the time parameter. A listener perceives a virtual source the direction of which is dependent on the gain factors.

2.1. Stereophony

Stereophonic listening configuration is most used listening setup. In it there are two loudspeakers placed in front of a listener, as illustrated in Fig. 2. The aperture of loudspeakers is typically $60°$. They are equidistant from the listener. In the figure variable θ denotes the perceived azimuth of the virtual source. There are a number of panning laws that estimate θ from the gain factors of loudspeakers. The estimated direction is called panning direction or panning angle.

Bennett et al. [17] derived a panning law for stereophonic setup by approximating the propagation path from ipsilateral loudspeaker to ear with a straight line, and the path from contralateral loudspeaker to ear with a curved line around the head. They ended up with a law that was earlier proposed for different listening conditions in [18]

$$\frac{\tan\theta}{\tan\theta_0} = \frac{g_1 - g_2}{g_1 + g_2}. \tag{2}$$

This equation is called the tangent law, the derivation of it is valid only with ITD when the frequency is below 500 Hz and when the listener's head is pointing directly between the loudspeakers. However, in [3] it is found that it is valid up to 1100 Hz with ITD cue, and also above 3000 Hz with ILD cue. In the equation it is also assumed that the elevation is $0°$. In [4] the tangent law has been formulated with θ_{cc} directions, which also allow one to use the laws when the ϕ_{cc} directions of loudspeakers are not the same.

Gain factors cannot be resolved as such with tangent law. They only state the relation between gain factors. To be able to solve the gain factors, an equation can be stated that keeps perceived virtual source loudness constant

$$\sqrt[p]{\sum_{n=1}^{n=N} g_n^p} = 1. \tag{3}$$

Here p can be chosen differently, depending on listening room acoustics, and its value affects the loudness of the virtual source [19]. In anechoic listening the virtual source loudness is roughly equal when $p = 1$, and in real rooms with some reverberation the value is often set to $p = 2$. The first case preserves roughly the amplitude of virtual source signals in the ear canals, and the latter case, the energy of them.

2.2. Multi-loudspeaker setups

In 2-D loudspeaker setups all loudspeakers are in the same plane with a listener. Pair-wise amplitude panning [20] methods can be used in such loudspeaker systems. The sound signal is applied to two loudspeakers between which the panning direction lies. Different panning laws have been used with such setups, although they have not been derived for such cases. When the existing laws are used in these cases, the directions of virtual sources may deviate from intended direction. This was observed in [21], where the localization of virtual sources was studied with listening tests. The direction of a virtual source changed inside a loudspeaker pair towards the median plane, when the pair was more on the side of a listener. When the centroid of a pair was in a lateral direction ($\pm 90°$), it was not at all possible to create stable virtual sources between loudspeakers.

There has been no panning laws derived for loudspeaker setups unsymmetric with the median plane, although panning laws for symmetric setups are commonly used in such cases. The prediction for virtual source direction perception may be false with traditional laws. This study focuses on derivation of a panning law that would hold in general cases.

A three-dimensional loudspeaker setup denotes a setup in which all loudspeakers are not in the same plane with the listener. Typically this means that there are some elevated and/or lowered loudspeakers added to a horizontal loudspeaker setup. Triplet-wise panning can be used in such setups. In it, a sound signal is applied to a maximum of three loudspeakers at one time that form a triangle from the listener's view point.

2.3. Vector base amplitude panning

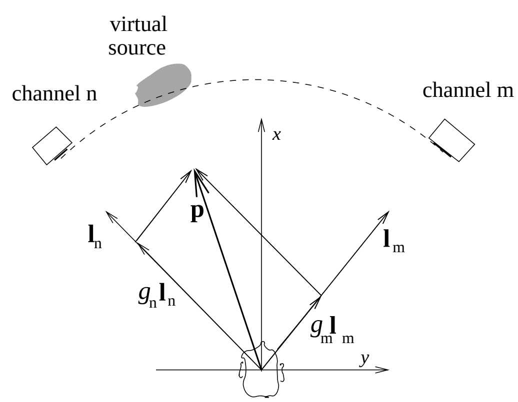

Figure 3: A loudspeaker pair for two-dimensional vector base amplitude panning (VBAP).

Vector base amplitude panning (VBAP) is a method to calculate gain factors for pair-wise or triplet-wise amplitude panning [5]. In pair-wise panning it is a vector reformulation of the tangent law (Eq. 2). Differing from the tangent law, it can be generalized easily for triplet-wise panning.

In VBAP the listening configuration is formulated with vectors; In two-dimensional panning a Cartesian unit vector $\mathbf{l}_n = [l_{n1}\ l_{n2}]^T$ points to the direction of loudspeaker n, from the listening position. Unit vectors $\mathbf{l}_n$, and $\mathbf{l}_m$ then define the directions of loudspeakers n, and m, respectively. The panning direction of a virtual source is defined as a 2-D unit vector $\mathbf{p} = [p_n\ p_m]^T$. A sample configuration is presented in Fig. 3.

The panning direction vector $\mathbf{p}$ is expressed as a linear combination of two loudspeaker vectors $\mathbf{l}_n$ and $\mathbf{l}_m$, and in matrix form:

$$\mathbf{p} = g_n\mathbf{l}_n + g_m\mathbf{l}_m, \tag{4}$$
$$\mathbf{p}^T = \mathbf{g}\mathbf{L}_{nm}. \tag{5}$$

Here g_n and g_m are gain factors, $\mathbf{g} = [g_n\ g_m]$ and $\mathbf{L}_{nm} = [\mathbf{l}_n\ \mathbf{l}_m]^T$. Vector $\mathbf{g}$ can be solved

$$\mathbf{g} = \mathbf{p}^T\mathbf{L}_{nm}^{-1} \tag{6}$$

if $\mathbf{L}_{nm}^{-1}$ exists, which is true if the vector base defined by $\mathbf{L}_{nm}$ spans a 2-D space. Eq. 6 calculates barycentric coordinates of vector $\mathbf{p}$ in a vector base defined by $\mathbf{L}_{nm}$. The components of vector $\mathbf{g}$ can be used as gain factors; a scaling of them may be desired according to Eq. 3.

3. LISTENING TESTS

The displacement of virtual sources with different loudspeaker setups is now measured using listening tests. The selection of listening test signals is considered first. Most of sounds that are used in main applications of amplitude panning, like gaming, virtual reality and movies, are broad band. Thus broad-band signals were used with different temporal and spectral structure in conducted listening tests. The spectrograms of four test signals are presented in Fig. 4. Pink noise was selected since it is considered to have a perceptually flat spectrum, and it has a quite strong temporal structure. It should thus produce reliable ITD and ILD cues. Chimes sound was selected since most of its energy is found at frequencies where ILD cue is strong. Speech is very important type of sound to humans, a male speech sample 'the concept of' was chosen. The sample contains different vowels, plosives, fricatives and a nasal. A snare drum tremolo was selected to produce strong ITD cues.

The test method was method-of-adjustment [22], four listeners were asked to adjusted the gain ratio of loudspeakers to match the direction of a virtual source to the same direction with a real source. In the tests the loudspeakers for virtual source production were in directions $-30°$ and $30°$, gain factors g_1 and g_2 were assigned to these loudspeakers, respectively. The reference sources were in $\pm 15°$ and $0°$ of azimuth, as is shown in Fig. 5. In first test the listener faced towards $0°$ direction, and in second and third test towards $-30°$ and $-60°$, respectively.

The listeners were sitting in an anechoic chamber equidistant from the loudspeakers, and held a keyboard on their lap. The keyboard was used to control the gain ratio between the loudspeakers. A virtual and a real source were produced consecutively and repeatedly, until the listener found a match between the sources and pressed a key. The initial gain ratio was random, and it was changed on a scale defined by the tangent law with one degree steps. The samples were presented in random order.

3.1. Results

The results are shown in Fig. 6 as a boxplot. They are discussed separately for different listener directions. With listener direction $0°$ the results are consistent; the lis-

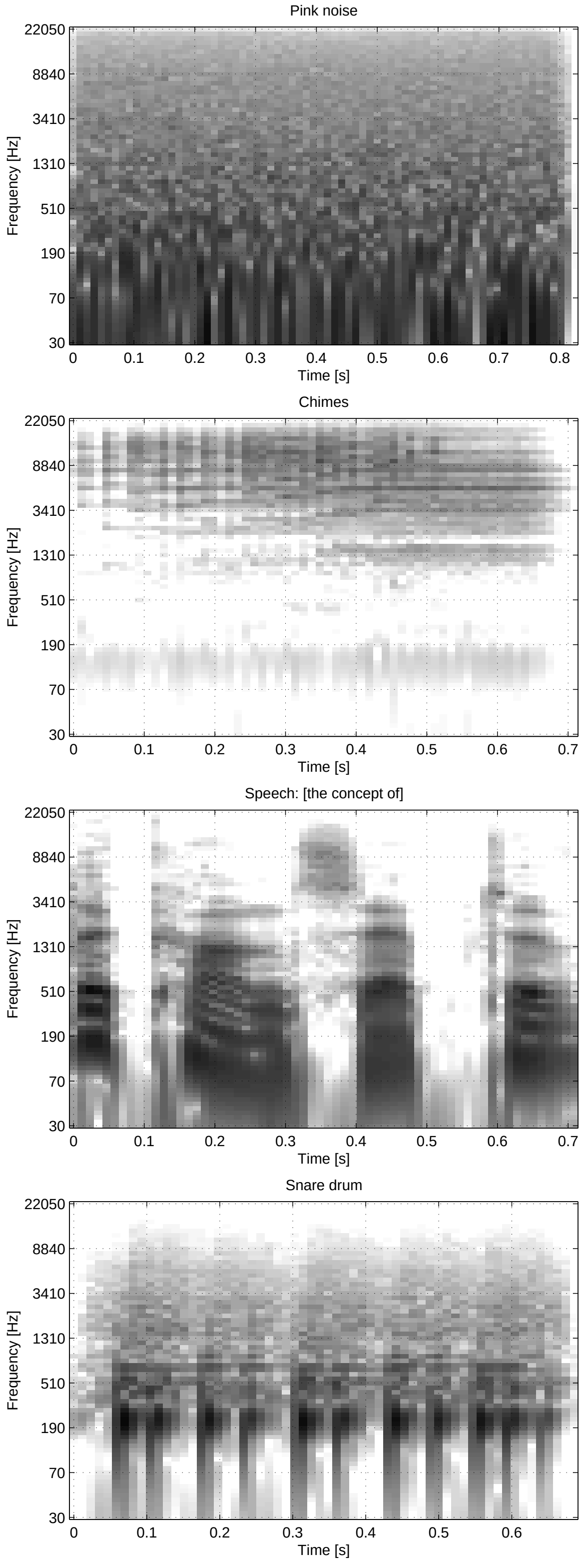

Figure 4: Spectrograms of listening test sound signals.

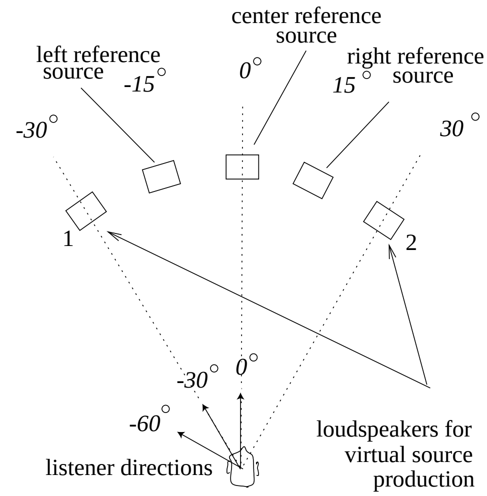

Figure 5: Listening test setup.

teners have favored certain gain ratios with different reference directions. The median values for different signal types are also near each other, however, the 'chimes' sound has produced somewhat different results. In the right side of the figure the VBAP panning directions that corresponds to gain ratios are shown. VBAP panning directions match well with the reference directions.

In [3] the ILD and ITD cues have been analyzed for virtual sources in this setup. The low-frequency ITD cues are stable and coincide with the panning direction. This explains the consistent result with signals that included low-frequency components (noise, speech and drum). The chimes sound was, however, adjusted differently. Simulation results in [3] show that ILD cues between 1 and 3 kHz suggest directions farther from the median plane when the panning angle is $\pm 15°$. This explains at least partly why the listeners favored panning directions towards the median plane in $\pm 15°$ reference case. However, the favored gain ratios of different signal types are relatively close to each other. It seems that the direction perception is similar with all tested signal types.

In $-30°$ listener direction test the adjusted gain ratios are near each other with each reference direction. However, the effect of sound signal is different from $0°$ listener direction case. It has an effect only with reference direction $-15°$. With reference directions $0°$ and $15°$ the sound signal seems not to have an effect on results.

The best judged gain ratios of listener direction $30°$ are displaced from gain ratios of listener direction $0°$ case with $2 - 5$ dB to negative direction. This is in agreement with earlier results in which a virtual source was found to be biased towards the median plane, when loudspeakers

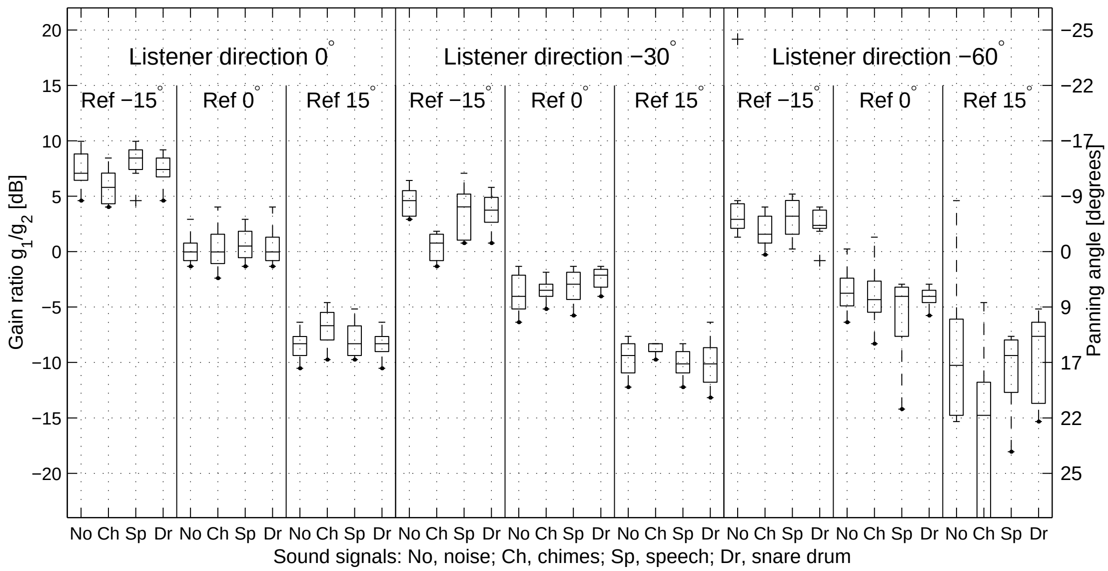

Figure 6: Listening test results. Best judged gain ratios of loudspeakers 1 ($-30°$) and 2 ($30°$). Four listeners matched a virtual source to same direction with a reference real source in direction $-15°$, $0°$ or $15°$. Two repetitions were conducted per each listener direction, each signal and each reference direction.

were moved to side of a listener [21]. The listeners have compensated the bias by making the loudspeaker louder that is farther from the median plane. This was also verified with an auditory model in [4]. ITD and ILD suggest directions biased towards the median plane when compared with VBAP panning direction.

In listener direction $-60°$ case, results are less stable with reference directions $0°$ and $15°$ than in previous tests. However, the results for reference direction $-15°$ are still quite stable. Two possible reasons can be stated for deteriorating of results. With reference directions $0°$ and $15°$ sources were in θ_{cc} direction $60°$ and $75°$ relative to listener's median plane. The accuracy of directional hearing is low at such lateral directions, and the quality of amplitude-panned virtual sources may be also low in these directions.

The best judged gain ratios are biased slightly more to negative values when compared with listener direction $-30°$ and $0°$. Corresponding VBAP panning angles are displaced slightly more than with $-30°$ listener direction case, which coincides with results in [21] and in [4]. There seems to be no strong influence of sound signal with $-60°$ head direction.

In overall, the effect of sound signal to best judged gain ratio is relatively small in all cases. This suggests that the panning law need not to be changed when sound signal is changed. However, the effect of listener direction is more prominent and systematic.

4. ENHANCING AMPLITUDE PANNING

Based on the results of previous chapter, it is clear that amplitude-panned virtual sources are displaced from their desired direction. A panning method which would produce less erroneous perceptions is now derived. The properties of a such panning law can be described as follows.

When a stereophonic loudspeaker pair is positioned symmetrically with the median plane, the tangent law (Eq. 2) predicts virtual source direction as well as possible with amplitude panning. The new panning law should be equivalent to the tangent law in such cases. When one loudspeaker is closer to the median plane than the other, virtual source is biased towards the median plane. A straight-forward way to compensate this is to make the level of loudspeaker higher which is farther away from the median plane to compensate the virtual source displacement, as the listeners did in reported listening tests. This can be implemented by adding an additional gain to each loudspeaker channel before gain factor normalization.

When direction $\pm 90°$ is between loudspeakers, there exists region between the loudspeaker where a virtual source can not be positioned. When the loudspeakers are positioned in a same cone of confusion, (e.g. $60°$ and $120°$ of azimuth), no stable virtual sources can be positioned between them. In these lateral cases the enhancement of amplitude panning is not possible. However, to be consistent with the tangent law, the enhanced panning law should be as similar as possible to it.

A enhancement of VBAP is presented in following section, that follows these guidelines.

4.1. Non-unitary VBAP

A panning method can be enhanced to compensate virtual source displacement by adding additional gain values after gain factor calculation, in pair-wise panning this can be formulated as

$$g_n^* = \frac{g_n}{\lambda_n}, \quad g_m^* = \frac{g_m}{\lambda_m}, \tag{7}$$

where λ_x is a positive real number related to corresponding loudspeaker channel. The g^* gain factors should be normalized using Eq. 3 to keep virtual source loudness constant. The normalization does not, however, change the ratio of gain factors.

In implementation of panning Eq. 7 should be computed during run time, that reserves computational capacity. We may seek a more efficient way to perform the enhancement. Using definitions of VBAP (Eqs. 4 – 6), Eq. 7 can be transformed to a form

$$[g_n^* \lambda_n \; g_m^* \lambda_m] = \mathbf{p} \mathbf{L_{nm}}^{-1}, \tag{8}$$

which can be expressed as

$$\mathbf{p} = g_n^* \mathbf{l}_n \lambda_n + g_m^* \mathbf{l}_m \lambda_m, \tag{9}$$

and further

$$\mathbf{g}^* = \mathbf{p}^T \mathbf{\Lambda}_{nm}^{-1} \tag{10}$$

where $\mathbf{g}^* = [g_n^* \; g_m^*]$ and $\mathbf{\Lambda} = [\mathbf{l}_n \lambda_n \quad \mathbf{l}_m \lambda_m]^T$ This formulation can be treated as a modification of VBAP (Eq. 4) in which the vectors in vector base are not of unitary length. Variable λ_x introduces a length to each unitary loudspeaker vector $\mathbf{l}_x$. This is called as non-unitary vector base amplitude panning (NVBAP). As shown, this is equivalent to Eq. 7. However, when formulated in this way, additional computational routines appear in loudspeaker setup initialization before run time, which ensures that the computational complexity of NVBAP is the same than of VBAP.

With a loudspeaker pair, standard VBAP produces gain factors that change symmetrically with the center point between loudspeakers, as presented in Fig. 7. The gain factors calculated with NVBAP for stereophonic setup with two different cases are shown in Fig. 7. In first case $\lambda_n = 2.0$ and $\lambda_m = 1.0$, in second case $\lambda_n = 4.0$ and $\lambda_m = 1.0$. It can be seen that the gain factor ratios are displaced with each other with a constant amount of decibels. This shows that NVBAP changes the ratio of loudspeaker gains, however, it is not known in this point how λ values should be set for loudspeaker pairs in different configurations. This is considered in next section.

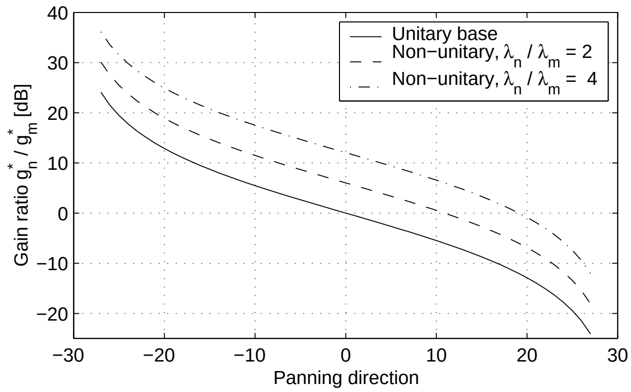

Figure 7: Gain ratio g_n/g_m when gain factors have been calculated with NVBAP using different λ_n/λ_m ratios. Loudspeakers n and m are positioned to $-30°$ and $30°$, respectively.

4.2. Defining compensation function

A function f of loudspeaker direction may be formed that produces λ values that compensate the displacement of amplitude-panned virtual sources with loudspeaker setups in lateral directions. In beginning of this section, some prerequisites were stated for the function. The effect of this enhancement should be similar in front and back of the listener, also, when direction $(90°, 0°)$ is inside a loudspeaker pair or triple, the effect should be mild or non-existing. These conditions are fulfilled when the function is defined depending on θ_{cc},

$$\lambda = f(\theta_{\text{cc}}). \tag{11}$$

This yields equal λ value within each cone of confusion, which also implies front-back symmetry. In stereophonic setup, and in all setups having two loudspeakers symmetrically with respect to the median plane, the function should return equal values, thus

$$f(\theta_{\text{cc}}) = f(-\theta_{\text{cc}}). \tag{12}$$

This yields left-right symmetry. To be consistent with the tangent law, it is defined that $f(\theta_{\text{cc}} = \pm 30°) = 1$. Since the virtual sources are displaced towards the median plane with lateral loudspeaker setups, the loudspeaker gains should be greater in those directions correspondingly. This can be achieved if the values of $f(\theta_{\text{cc}})$ are descending when θ_{cc} departs from zero. Defining $f(\theta_{\text{cc}})$ is discussed in next section.

Experimental data from listening tests and simulations is analyzed to define the compensation function $f(\theta_{\text{cc}})$. To be able define it, it is desired to know how much the ratio between two loudspeaker gains should be changed with loudspeaker sets in different directions.

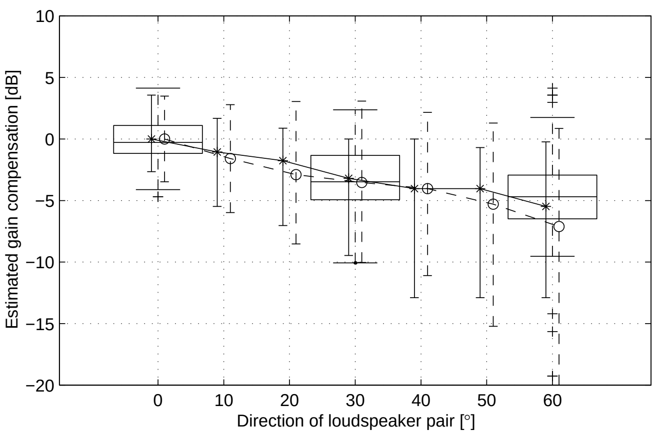

Figure 8: Boxplot: average gain change needed to compensate virtual source displacement in listening test data. Lines: average gain change needed to compensate auditory cue values, Stars and solid line, median of simulated gain changes for ITD, Circles and dashed line, median of simulated gain changes for ILD. The whiskers denote the quartiles of data.

Gain ratios are known, that produced to the listening test attendees a perception of a virtual source in the same direction with a reference source. These gain ratios are compared to the gain ratios that VBAP law suggest. Both ratios are measured in decibels and a difference is taken. This estimates how much the gain factors should be changed to enhance the panning law. The differences calculated with listening test data are shown as boxplots in Fig. 8. The medians estimate the λ_1/λ_2 ratios with each used loudspeaker setup.

However, the listening test results provide data only with three different loudspeaker setups. To obtain more data, the auditory model described in Sec. 1.3 was used to calculate auditory cues with different loudspeaker setups. Virtual sources were simulated using a loudspeaker pair with aperture $60°$ with centroid in directions $0°, 10°, ..., 60°$. In each centroid direction virtual sources were simulated with $10°$ intervals for 6 individuals. A set of ILDA:s and ITDA:s was obtained, and a difference between desired value (panning direction) and cue values was computed. This estimates the directional displacement of virtual sources. The amount of gain factor compensation can be estimated with VBAP. It is assumed that if a virtual source is displaced by x degrees, it can be compensated with $-x$ degrees change in panning direction, which yields the needed gain ratio change. The data is shown also in Fig. 8.

The gain compensation estimates calculated from listening test and auditory modeling results presented in Fig. 8 correspond well with each other. The medians of auditory modeling results are close to the medians of listening test results, however, with listener direction $-60°$ the auditory model results deviate from listening test data slightly. This suggests that the auditory modeling can be used also with listener directions that were not included in the listening tests, which provides more data for deriving the compensation function.

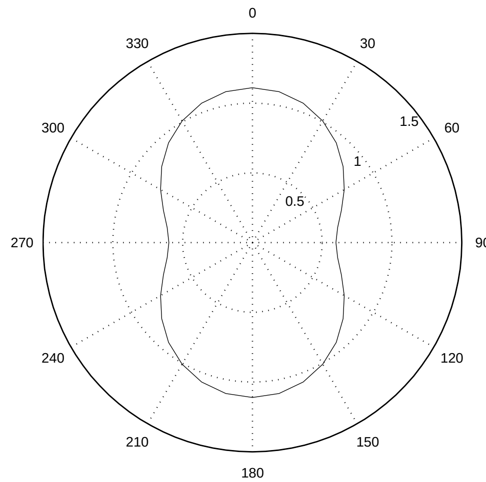

Figure 9: Compensation function. The dependence of λ on loudspeaker direction.

A compensation function is now derived. A first-order spherical harmonic

$$f(\theta_{\mathrm{cc}}) = a\sin\theta_{\mathrm{cc}} + b\cos\theta_{\mathrm{cc}}, \tag{13}$$

where a and b are constants, fulfills the prerequisites of symmetry suggested in Sec. 4.1. The constants are selected in a way that λ_1/λ_2 ratio matches to results presented in Fig. 8. The values were searched manually, and a quite good fit with the results was found with values $a = 1.11$ and $b = 0.6$. The compensation function is shown in Fig. 9.

To verify the method, virtual sources were simulated with NVBAP panning angle similarly as earlier in this section. The differences between auditory cue values and panning directions were taken, and resulting deviations are shown in Fig. 10 for loudspeaker apertures $60°$ and $30°$. The bias of both cues is non-existent with loudspeaker set directions $0°$ to $40°$. With set directions $50°$ and $60°$ the medians deviate from $0°$, especially with ILD cues. However, the target direction $0°$ lies still between quartiles of data. This suggests that NVBAP compensates the systematic bias.

The results of listening test conducted in this study can also be used to verify the defined compensation function. In the tests, the attendees selected a gain ratio g_1/g_2 that created a virtual source in direction of a reference

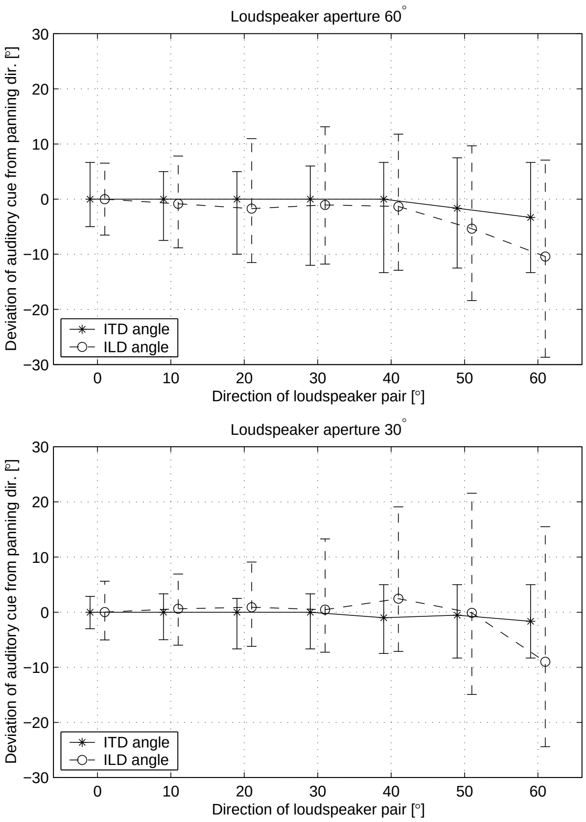

Figure 10: Auditory cues of a loudspeaker pair with $60°$ and $30°$ apertures in different directions. Virtual sources were simulated within each pair at intervals of $10°$. The difference of panning direction and auditory cue angle was calculated with each cue and each frequency. Medians of ITDA and ILDA deviations are shown, the whiskers denote the quartiles of the data.

source. If the best matched gain ratios would correspond with NVBAP panning angles, it would support the hypothesis that NVBAP compensates the virtual source displacement. The preferred gain ratios were converted to corresponding panning directions using NVBAP (Eq. 9) with presented compensation function. The listening test data as resulting NVBAP directions is presented in Fig. 11. Panning angles correspond well to reference source directions in most cases, and there exists no systematic bias towards or away from the median plane.

5. CONCLUSIONS

In pair-wise panning, when loudspeakers are not symmetrically with the median plane of a listener, amplitude-panned virtual sources are biased towards the median plane when compared with VBAP panning direction. The displacement was measured with listening tests. It was found to be of order $10°$ in some cases, which may be disturbing.

The displacement can be compensated by changing the ratio of gain factors by favoring the loudspeaker that is farther away from the median plane. A non-unitary vector base amplitude panning (NVBAP) method was derived, that compensates the displacement. A function that defines the amount of compensation was formed using listening test results and results from a binaural auditory model.

The listening test results were interpreted using NVBAP with the compensation function. This showed that the virtual sources were not biased towards the median plane when compared with NVBAP panning direction. Virtual sources created with NVBAP were also simulated using loudspeaker pairs with a $30°$ and $60°$ aperture. The simulation suggested that the bias is also compensated in these cases.

ACKNOWLEDGMENT

The work of Dr. Pulkki has been supported by the Graduate School in Electronics, Telecommunications and Automation (GETA) of the Academy of Finland. The author also wishes to thank all listening test attendees.

REFERENCES

[1] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. Creating interactive virtual acoustic environments. *J. Audio Eng. Soc.*, 47(9):675–705, Sept 1999.

[2] H. Fouad, J. A. Ballas, and D. Brock. An extensible toolkit for creating virtual sonic environments. In *Proceedings of International conference on auditory displays 2000*, Atlanta, Georgia, USA, 2000. ICAD.

[3] V. Pulkki and M. Karjalainen. Localization of amplitude-panned virtual sources I: Sterephonic panning. a companion paper submitted to J. Audio Eng. Soc., 2001.

[4] V. Pulkki. Localization of amplitude-panned virtual sources II: three-dimensional panning. submitted to J. Audio Eng. Soc., 2001.

[5] V. Pulkki. Virtual source positioning using vector base amplitude panning. *J. Audio Eng. Soc.*, 45(6):456–466, June 1997.

[6] J. Blauert. *Spatial Hearing, Revised edition*. The MIT Press, Cambridge, MA, USA, 1997.

[7] Richard O. Duda. Elevation dependence of the interaural transfer function. In Robert H. Gilkey and Timothy R. Anderson, editors, *Binaural and*

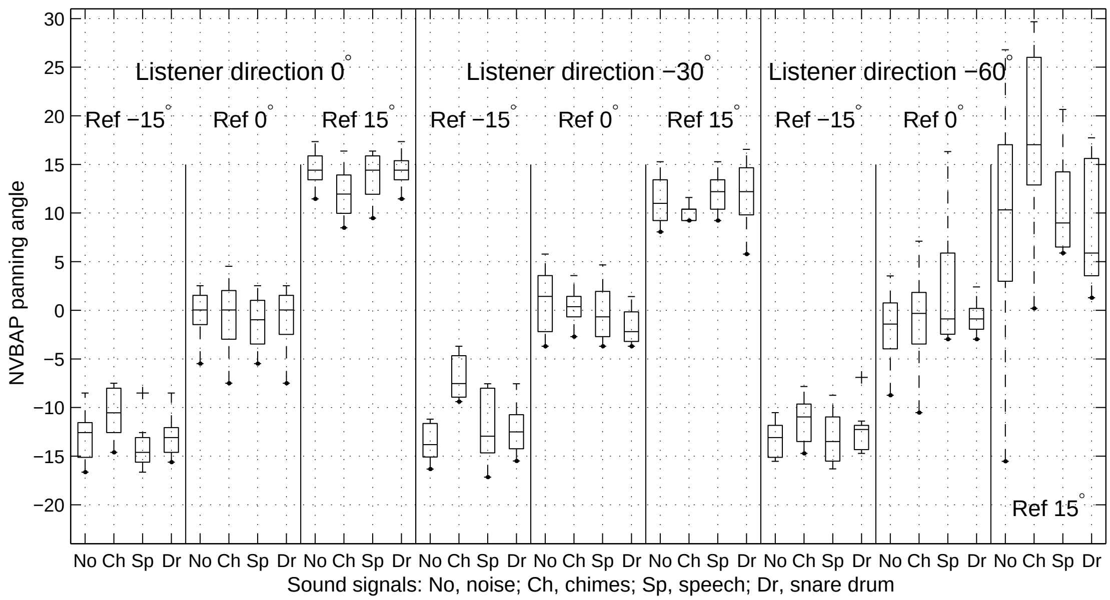

Figure 11: Best matched gain ratios in listening tests (Fig. 6 being transformed to NVBAP panning angles.)

Spatial Hearing in Real and Virtual Environments, pages 49–75. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997.

[8] R. H. Gilkey and T. R. Anderson, editors. *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum Assoc., Mahwah, NJ, US, 1997.

[9] D. S. Brungart and W. M. Rabinowitz. Auditory localization of nearby sources. i: Head-related transfer functions. *J. Acoust. Soc. Am.*, 106(3):1465–1479, 1999.

[10] P. M. Zurek. The precedence effect. In W. A. Yost and G. Gourewitch, editors, *Directional Hearing*, pages 3–25. Springer-Verlag, 1987.

[11] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Gutman. The precedence effect. *J. Acoust. Soc. Am.*, 106(4):1633–1654, Oct. 1999.

[12] B. C. J. Moore. *An introduction to the psychology of hearing*. Academic Press, San Diego, fourth edition, 1997.

[13] R. Patterson, K. Robinson, J. Holdsworth, D. Mckeown, C. Zhang, and M. H. Allerhand. Complex sounds and auditory images. In L. Demany Y. Cazals and K. Horner, editors, *Auditory Physiology and Perception*, pages 429–446. Pergamon, Oxford, 1992.

[14] L. A. Jeffress. A place theory of sound localization. *J. Comp. Physiol. Psych.*, 61:468–486, 1948.

[15] R. M. Stern and C. Trahiotis. Models of binaural perception. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 499–531. Lawrence Erlbaum Assoc., Mahwah, NJ, USA, 1997.

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

[17] J. C. Bennett, K. Barker, and F. O. Edeko. A new approach to the assessment of stereophonic sound system performance. *J. Audio Eng. Soc.*, 33(5):314–321, May 1985.

[18] D. M. Leakey. Some measurements on the effect of interchannel intensity and time difference in two channel sound systems. *J. Acoust. Soc. Am.*, 31(7):977–986, July 1959.

[19] F. R. Moore. *Elements of Computer Music*. Prentice Hall, Englewood Cliffs, New Jersey 07632, 1990.

[20] J. Chowning. The simulation of moving sound sources. *J. Audio Eng. Soc.*, 19(1):2–6, 1971.

[21] G. Theile and G. Plenge. Localization of lateral phantom sources. *J. Audio Eng. Soc.*, 25(4):196–200, April 1977.

[22] B. L. Cardozo. Adjusting the method of adjustment: SD vs DL. *J. Acoust. Soc. Am.*, 37(5):768–792, May 1965.