DIRECTIONAL QUALITY OF 3-D AMPLITUDE PANNED VIRTUAL SOURCES

Ville Pulkki, Matti Karjalainen

Helsinki University of Technology
Acoustics Laboratory
P.O.Box 3000, FIN-02015 HUT, Finland
Ville.Pulkki@hut.fi, Matti.Karjalainen@hut.fi

ABSTRACT

When a listener is surrounded by multiple loudspeakers, some of which are located above and/or below him/her, an immersive three-dimensional auditory display can be created with vector base amplitude panning (VBAP). In VBAP a complex loudspeaker layout is divided to loudspeaker triplets or pairs (sets). A sound signal is applied to loudspeakers of one set with different gain factors at one time, which creates a virtual source to a direction that does not necessarily correspond with any of the loudspeakers. The directional quality of virtual sources created with VBAP is studied in this paper. Listening tests are conducted with loudspeaker pairs in horizontal and vertical layouts, and with a loudspeaker triplet. The results are interpreted with a binaural model. The binaural model is further used to predict directional quality of virtual sources in arbitrary loudspeaker setups. It is found that VBAP predicts the angle between the median plane and a perceived virtual source quite accurately when the loudspeakers of a set are near the median plane. When loudspeakers are moved to a side of the listener, virtual source direction is biased towards the median plane. The elevation of virtual sources is perceived individually and can not be predicted with any panning law.

1. INTRODUCTION

In amplitude panning a same sound signal is applied to a number of loudspeakers in different directions equidistant from the listener. A virtual source appears to a direction that is dependent on amplitudes of the loudspeakers. The direction may not coincide with any physical sound source. Most typically amplitude panning has been used with stereophonic loudspeaker setup. However, it is increasingly used to position virtual sources to arbitrary loudspeaker setups. Vector base amplitude panning (VBAP) [3] is a method to position virtual sources to such loudspeaker setups. Although VBAP has been formulated for arbitrary loudspeaker setups, it has not been known what is the quality of generated virtual sources.

The perceptual quality of virtual sources is known quite well in horizontal two-loudspeaker setups, especially in the stereophonic setup [4]. However, the localization mechanism has been explained only at low frequencies. The reasons why high frequencies are localized consistently with low frequencies has not been explained. In setups in which the loudspeakers are not located in a same elevation, the quality of virtual sources remains unexplored. The results of horizontal setups may not be valid, since the localization may not be based on same mechanisms.

In this paper the directional quality of amplitude-panned virtual sources is discussed. Results are gathered from listening tests and simulation results conducted in this work and earlier.

2. SPATIAL HEARING

2.1. Theory and coodinate systems

The duplex theory of sound localization states that the two frequency-dependent main cues of sound source localization are the *interaural time difference* (ITD) and the *interaural level difference* (ILD). They are caused by the wave propagation time difference (primarily below 1.5 kHz) and the shadowing effect of the head (primarily above 1.5 kHz), respectively [4, 5]. The cues include information in which cone of confusion the sound source is located. A cone of confusion can be approximated with a cone which has a symmetry axis along a line passing through the listener's ears, as picted in Fig. 1. The direction perception inside a cone of confusion is refined using other cues.

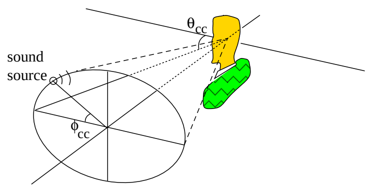

Figure 1: Cone of confusion. Spherical coordinate system convenient in direction hearing.

Conventionally, the location of sound sources is denoted with azimuth (θ) and elevation (ϕ) angles. If elevated sound sources are also applied in auditory research, this formulation is inconvenient, since a cone of confusion cannot be easily formulated. An alternative spherical coordinate system has been used by Duda [6]. The sound source location is specified by defining the cone of confusion in which it lies, and further by the direction within the cone of confusion. The cone of confusion is defined by an angle θ_{cc} between the median plane and the cone, as in 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 $180°$. These coordinates are used throughout this paper. Standard nomenclature for these variables is lacking, though it would be beneficial. In some studies θ_{cc} is referred to as the left/right (L/R) direction.

The perception of θ_{cc} relies therefore on ITD and ILD decoding most, however, the perception of ϕ_{cc} is not understood thoroughly. It is known that the information is decoded from spectral contents of monaural signals [7], however, the cues that are used in spectrum are not known in detail.

2.2. Evaluation of directional quality of virtual sources

The directional quality of virtual sources is defined to describe how well the perceived direction of a virtual source corresponds to the direction where it was intended to be. If a point-like virtual source is produced to an intended direction, perfect quality is achieved. In a such case the auditory cues of the virtual source would be identical with a corresponding real source. Typically there are some deviations in cues of a virtual source from the cues of real source. The cues may propose different directions with different frequencies and cues. It may then be perceived to be spatially spread or diffuse. The directional quality of a virtual source may be estimated computationally by calculating the frequency-dependent ILD and ITD cues and by comparing them with real source cues.

2.3. Modeling the direction perception of virtual sources

To simulate the perception of virtual sources, a binaural auditory model is used to calculate localization cues for signals arriving at the ear canals. The model of auditory localization is presented in Figs. 2 and 3, and it consists of the following parts:

- simulation of ear canal signals using measured HRTFs
- a binaural model of neural decoding of ITD and ILD cues consisting of
  - middle ear model
  - cochlea model with 42 ERB bands
  - model of ITD decoding
  - model of ILD decoding
  - model of loudness level spectrum decoding
- a model of high-level perceptual processing for the θ_{cc} direction decoding based on a data-base search

The model estimates the ITD angle (ITDA) and ILD angle (ILDA) as functions of the equivalent rectangular bandwidth (ERB) pitch scale. The ITDA and ILDA present θ_{cc} directions where a real source should be placed to produce the virtual source ITD or ILD. Directional quality of amplitude-panned virtual sources is estimated by calculating deviations between panning direction and virtual source cue angles at different frequencies and with different cues. These models are explained in [8, 9, 10]. ITD cue is stable with frequency and monotonic with θ_{cc}, thus an ITDA estimation can be assumed to be valid at all frequencies and at most θ_{cc} directions. 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 $-\gamma < \theta_{cc} < \gamma$ [4] depending on frequency, where the value of γ is typically $40° - 80°$. Thus ILDA cue may not carry information of source location far away from the median plane. Additionally, ILD may be dependent also on ϕ_{cc}.

3. VECTOR BASE AMPLITUDE PANNING

In VBAP a loudspeaker triplet or pair is formulated with vectors, as presented in Fig. 4. The unit-length vectors $\mathbf{l}_m, \mathbf{l}_n$ and $\mathbf{l}_k$ point from listening position to the loudspeakers. The panning direction is presented with unit-length vector $\mathbf{p}$. Vector $\mathbf{p}$ is expressed as a linear weighted sum of the loudspeaker vectors

$$\mathbf{p} = g_m\mathbf{l}_m + g_n\mathbf{l}_n + g_k\mathbf{l}_k. \quad (1)$$

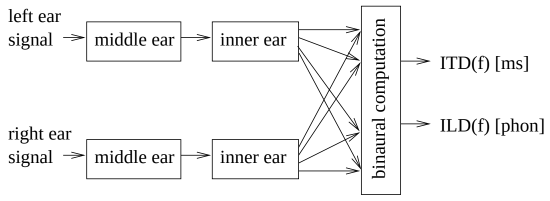

Figure 2: Binaural model of directional cue decoding.

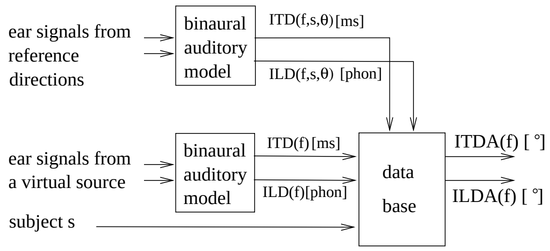

Figure 3: Functional model of auditory localization.

Here g_m, g_n, and g_k are called gain factors of respective loudspeakers. The gain factors can be solved as $\mathbf{g} = \mathbf{p}^T\mathbf{L}_{mnk}^{-1}$, where $\mathbf{g} = [g_m \; g_n \; g_k]^T$ and $\mathbf{L}_{mnk} = [\mathbf{l}_m \; \mathbf{l}_n \; \mathbf{l}_k]$. The calculated factors are used in amplitude panning as gain factors of the signals applied to respective loudspeakers after suitable normalization, e.g. $||\mathbf{g}|| = 1$. VBAP can be formulated for loudspeaker pairs easily, the amount of vectors in the vector base is then two.

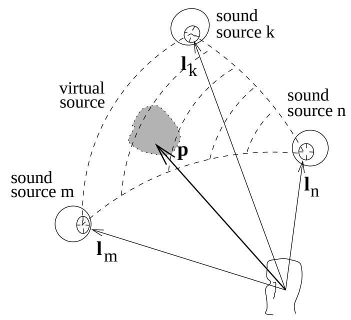

Figure 4: A loudspeaker triplet for three-dimensional vector base amplitude panning (VBAP).

4. LISTENING TESTS

In this section listening tests that have been conducted are described. The tests are described more detailed in [9, 10]. The listeners adjusted the panning angle of a virtual source to match with a real source in a large anechoic chamber. The number of loudspeakers was two or three, and they were set up in different manner. The sound signals were 1/3-octave band passed noise on ERB scale, or full band pink noise.

The listeners had normal hearing. They were sitting equidistant from the loudspeakers, and held a keyboard on their lap. The keyboard was used to control the panning angle. The virtual and the real source were produced consecutively and repeatedly until the listener found a match between the sources and pressed a key.

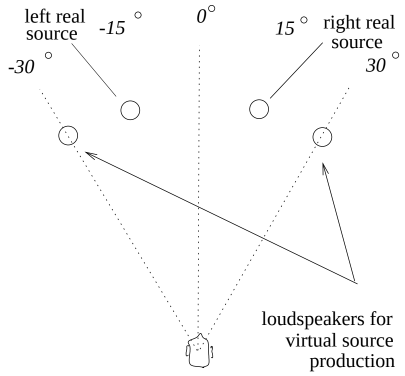

Figure 5: Horizontal listening test setup.

4.1. 2-D panning in the horizontal plane

In this test the loudspeakers were in stereophonic arrangement and the real sources were in $\pm 15^\circ$ of θ_{cc}, as is shown in Fig. 5. Eight listeners adjusted a virtual source containing 1/3-octave band noise at 11 frequency bands to match with real sources. The adjustment was conducted two times to both real sources, producing 4 best matched panning angles for each frequency band.

The polarities of panning angles with real source -15° were changed. The data was combined with 15° data forming a new data set. This reduces bias caused by unsymmetries in listening setup.

The obtained data was tested with appropriate non-parametric tests with respect to variables *frequency*, *subject* and *repetition*. It was found that only the effect of variable *frequency* was significant ($p < 0.01$). The judgments made by the subjects are shown as a function of frequency in Fig. 6. A prominent frequency-dependence can be seen. The data has in most cases a value between $12^\circ - 16^\circ$, but there exists a dip around 1 - 2 kHz.

This suggests that the virtual source with panning angle 15° is localized to 15° at low and high frequencies, but near 1.7 kHz there are some irregularities in the behaviour. The fact that listeners favored panning angles with smaller absolute value proposes that the virtual sources would be perceived to larger absolute values than 15° with $\pm 15^\circ$ panning angle.

To explore this, a virtual source panned to 15° was simulated with the auditory model. The resulting cues are shown in Fig. 7. Low-frequency ITDAs and high-frequency ILDAs coincide with panning angle values, whereas near 1.7 kHz there is a large bump in values, especially with ILDA values. The frequency-dependence found in listening test results can be explained: the dip in listening tests corresponds to the bump in simulation results. The listeners have been compensating high values of cues when they were panning narrow-band sources near 1.7 kHz. This also explains why high frequencies are localized consistently with low frequencies; the high-frequency ILDA cue coincides with low-frequency ITDA cue roughly.

4.2. 2-D panning in the median plane

Two loudspeakers for virtual source generation were positioned to the median plane in front of the listener in ϕ_{cc} directions -15° and 30° as shown in Fig. 8. The real sources were in 0° and 15°

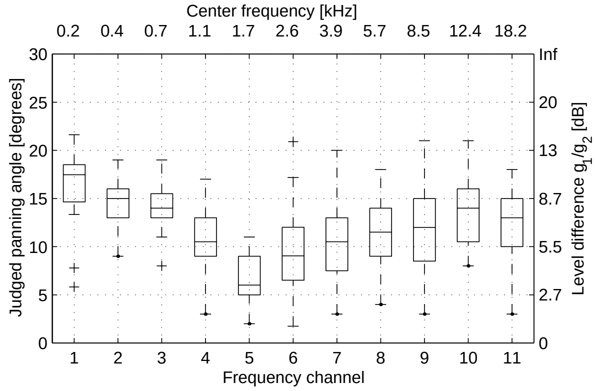

Figure 6: Best judged panning angles with 1/3-octave noise in stereophonic listening. Real source $\pm 15^\circ$, head forward, 8 subjects. The adjustments were conducted to both reference real sources twice.

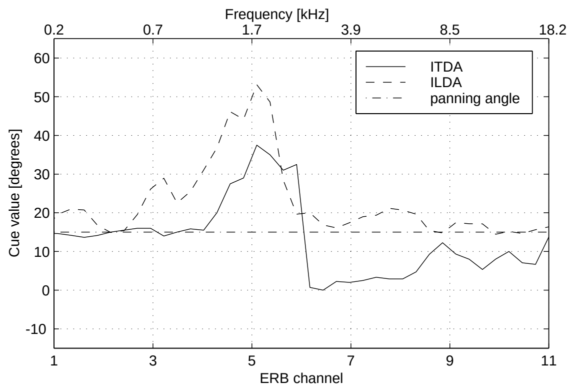

Figure 7: Simulated auditory cues of an amplitude-panned virtual source with 15° panning angle in stereophonic listening. Cue angle values calculated were with 20 individual HRTF sets. The median of resulting values is shown.

directions. The listeners adjusted a virtual source with broad band noise four times to each of real sources.

The data was tested to fulfill assumptions for analysis of variance (ANOVA) of normal distribution of data and residuals. The ANOVA test was run to find the effects of the following variables: *subject*, *source* (0° or 15° of ϕ_{cc}) and *repetition*. Also, all two-way interactions were tested.

The variable *repetition* did not significantly affect the results, neither its interaction with *subject* or *source* were significant ($p > 0.05$). This implies that learning was not a significant factor in judgments.

The effect of the variable *source* was significant ($p < 0.001$). The results are presented for each reference real source and for each individual in Fig. 9. The responses for the reference real source at 15° have greater values than for the reference real source at 0°. The listeners thus had to adjust the panning angle to larger values to perceive the virtual source in higher ϕ_{cc} directions. This suggests that the perceived ϕ_{cc} direction varies monotonically with the ϕ_{cc} panning angle.

The effect of the variable *subject* and the effect of the interac-

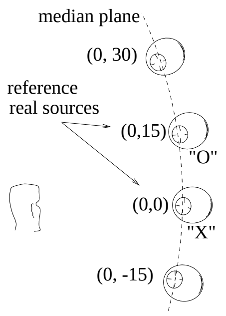

Figure 8: Setup used in the median plane listening tests. Directions expressed as (θ_{cc}, ϕ_{cc}) pairs.

tion *subject* * *source* were significant ($p < 0.001$). These features suggest that each subject needs an individual panning angle to perceive the virtual source in the ϕ_{cc} direction of each reference real source. Based on this result it is clear that the ϕ_{cc} direction perception of an amplitude-panned virtual source is individual, which yields that a single law that would predict virtual source direction correctly for all individuals cannot be formed.

The responses of individual subjects are now discussed. In Fig. 9 it can be seen that subjects other than NP, LS, and TL have panned the virtual sources in a relatively consistent way. Here, to be consistent means that the virtual sources matched to the $0°$ reference real source have a lower ϕ_{cc} panning angle than the virtual sources matched to the $15°$ reference real source. Also, the panning angle distributions of each individual to each reference real source are relatively narrow. This suggests that the perception of the virtual source is similar each time, and that they have not been guessing in the test.

Subjects NP, LS and TL have panned virtual sources corresponding to different reference real sources to almost the same panning angle values. This suggests that these subjects perceived the virtual source to jump from $-15°$ to $30°$ of ϕ_{cc} with very few panning steps.

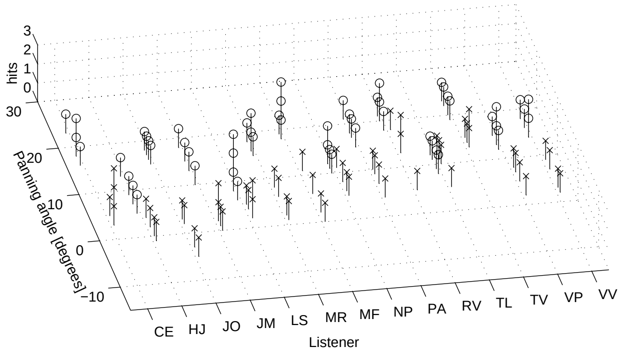

Figure 9: Best judged panning angles in the median plane listening test with pink noise. Crosses indicate the panning angles for reference real source at a $0°$ ϕ_{cc} direction, and circles indicate panning angles for the reference real source at a $15°$ ϕ_{cc} direction.

The individual monaural loudness level spectra of seven listeners were analyzed in [10]. However, the analysis failed, the performance of subjects could not be explained using current models of elevation perception.

4.3. 3-D panning

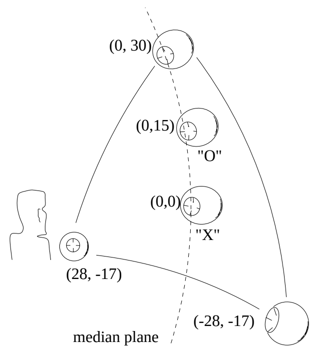

Figure 10: Loudspeaker setup for triplet listening test. Directions expressed as (θ_{cc}, ϕ_{cc}) pairs.

Eight listeners conducted adjustment with broad band noise to a loudspeaker triplet in front of the listener as shown in Fig. 10. The results of four listeners are shown in Fig. 11, the non-shown results of four listener resemble these results. It can be seen that the listeners adjusted the θ_{cc} direction consistently with themselves and with other attendees. The favored θ_{cc} also matches with prediction of panning laws. The favored ϕ_{cc} directions changes prominently between listeners, however, the listeners adjusted it consistently with themselves in general.

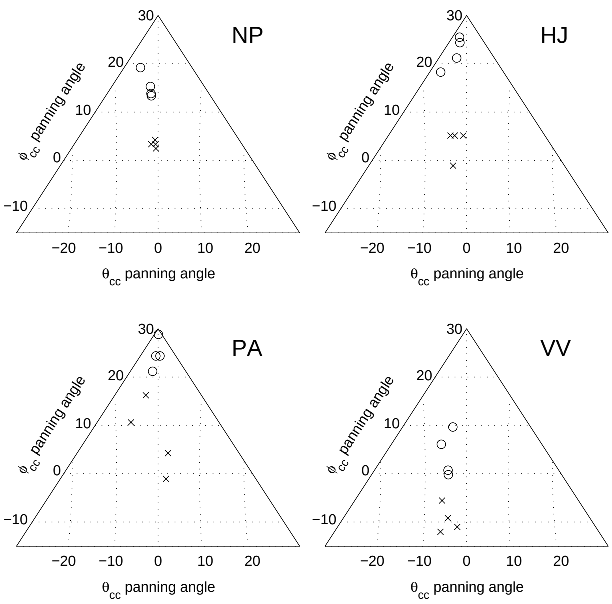

Figure 11: Best matched panning directions for triplet 1 with reference real sources in directions $(0°, 0°)$ (crosses) and $(0°, 15°)$ (circles).

This result seems to be a superposition of previous test results. Although the number of loudspeakers is now three instead of two, and they were arranged as a triangle, the θ_{cc} panning angle corresponds well to perceived direction, but ϕ_{cc} does generally not. Similar results were obtained in tests with a loudspeaker triangle which had loudspeakers in positions $(0°, 30°)$, $(0°, -15°)$ and $(-45°, -15°)$, the results are presented in [10].

5. MEASURING VIRTUAL SOURCE QUALITY WITH BINAURAL AUDITORY MODEL

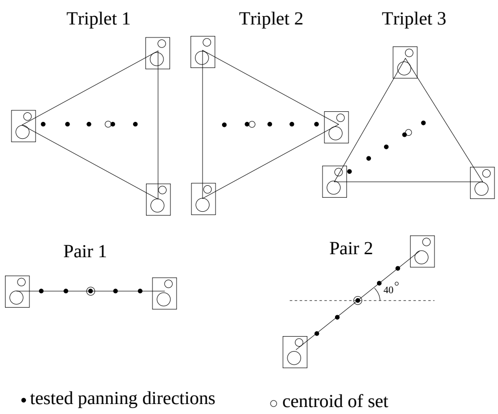

Figure 12: *Loudspeaker sets used in the simulation of virtual source θ_{cc} direction perception. Each solid line between two loudspeakers span a $60°$ angle from the listener's view point.*

To investigate the θ_{cc} direction perception more generally, model-based simulations with three triplets and two pairs were conducted. The loudspeaker sets and panning directions inside each set are shown in Fig. 12. The simulations were conducted with sets in four directions with respect to the listener. The centroids of the sets were set to θ_{cc} directions of $0°$, $30°$, $60°$, and $90°$.

The signals appearing at the subjects' ears were simulated with measured KEMAR HRTFs [11]. The ITDAs and ILDAs were simulated with the auditory model. The deviation between the panning angle and resulting ITDAs and ILDAs were calculated in all cases. The ITDA deviations below 1000 Hz and ILDA deviations at all frequencies are plotted in Fig. 13. The ITDA deviations above 1000 Hz are not presented since they behave mostly erroneously and in an unstable manner.

The ITDA simulations are discussed. When the centroid of panning sets is at $0°$ of θ_{cc}, the ITDA deviations are small regardless of the loudspeaker set. This could be assumed for pairs based on previous results, but it is a new result for triplets. The spread of triplet deviations are slightly larger, though. However, this suggests that the panning angle describes reasonably accurately the perception of the virtual source θ_{cc} direction also in triplet panning.

When the centroid of a set is at $90°$ of θ_{cc}, the loudspeakers of that set are all at $60°$ of θ_{cc}, i.e., they are thus in a same cone of confusion. Theoretically it has been proposed that the θ_{cc} directions of virtual sources are between the θ_{cc} values of the loudspeakers [10]. Thus the perceived θ_{cc} direction of virtual sources should have in these cases a value near $60°$, which should create a prominent negative deviation. Indeed, the average deviation is of

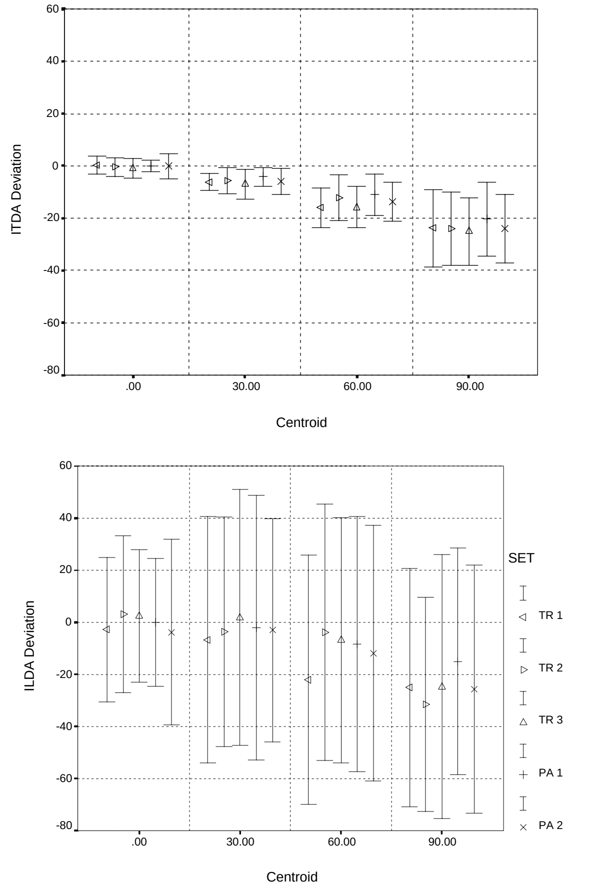

Figure 13: *θ_{cc} direction simulation with different setups. upper: Deviations between the θ_{cc} panning angle and ITDAs below 1000 Hz (12 bands). lower: Deviations between the θ_{cc} panning angle and ILDAs (42 bands). x-axis: the direction of the centroid of a loudspeaker set. A mean value is used as a measure of the centroid, and standard deviation is used as a measure of spread.*

the order of $-25°$, and follows the proposition. This seems to be valid both with pairs and triplets.

With a set centroid direction of $0°$, the negative bias phenomenon does not exist. When the set centroid direction increases, the negative bias increases as well. This explains the listening test results of Theile & Plenge [12], where it was found that virtual sources are localized nearer the median plane when a loudspeaker pair is moved towards the side of the subject.

The spread of ITDA deviations increases also with the set centroid θ_{cc} direction. To explore this, the ITDA deviations were plotted with each loudspeaker set direction with respect to frequency and panning angle. The plots (not shown here) implied that the deviation is not related to frequency, but the effect of the panning angle is prominent. This suggests that ITD cues of virtual sources are quite consistent, but are biased from the panning angle.

The ILDA deviations also have more negative values with increasing direction of the set centroid θ_{cc} direction. However, the ILDA deviations are spread out for all centroid directions prominently. When the source of the spread was explored, it was found that the ILDA deviation is dependent both on frequency and panning angle. This suggests that the ILDA cues of amplitude-panned virtual sources are not in general consistent. Only when the loudspeaker set is symmetric with the median plane, do the ILDA values have roughly correct values at high frequencies. However, the discrepancy between ILDA values and panning angle may be disregarded since it is known that ILD does not provide consistent cues even for real sources in lateral directions [4].

In the listening tests conducted in this paper it was found that the perception of the virtual source ϕ_{cc} direction is highly individual in the median plane. Since the localization is based on binaural or monaural spectral mechanisms also in other cones of confusion, it can be assumed that the perception of the amplitude-panned virtual source ϕ_{cc} direction is individual in all directions.

6. CONCLUSIONS

The panning direction describes quite accurately the perceived virtual source θ_{cc} direction with loudspeaker pairs and also with loudspeaker triplets, when the center of a pair or a triplet is near the median plane. The low-frequency ITD angle and roughly the high-frequency ILD angle coincide with the θ_{cc} panning direction.

If a pair or a triplet is to the left or right of a subject, the virtual source is located closer to the median plane than the θ_{cc} panning angle predicts, mostly due to low-frequency ITD behavior. ILD cues are generally distorted heavily with pairs or triplets in lateral directions.

The ϕ_{cc} directions of amplitude-panned virtual sources are perceived individually. Only if the loudspeakers share the same ϕ_{cc} direction is the virtual source localized mostly to that ϕ_{cc} direction. In other cases ϕ_{cc} direction perception cannot be predicted accurately. However, in conducted listening tests, most of the listeners could find ϕ_{cc} panning angles that created a virtual source direction perception to a reference real source. The ϕ_{cc} direction perception was also shown to be monotonic with the ϕ_{cc} panning angle.

The directional quality of amplitude-panned virtual sources seems to be twofold. The directional quality of virtual sources in θ_{cc} direction is fairly good, especially near the median plane. Amplitude panning seems to be unable to create similar ϕ_{cc} direction perceptions to different listeners, which makes quality worse on that side. However, the virtual source ϕ_{cc} direction is typically perceived to match with the loudspeakers that were used, and the accuracy of human spatial hearing quite bad in ϕ_{cc} direction perception. With a fairly dense loudspeaker array the quality of amplitude-panned virtual sources is therefore acceptable for many applications.

7. ACKNOWLEDGMENT

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

8. REFERENCES

[1] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 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, "Virtual source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, June 1997.

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

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

[6] Richard O. Duda, "Elevation dependence of the interaural transfer function," in *Binaural and Spatial Hearing in Real and Virtual Environments*, Robert H. Gilkey and Timothy R. Anderson, Eds., pp. 49–75. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997.

[7] F. L. Wightman and D. J. Kistler, "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds. Lawrence Erlbaum Assoc., Mahwah, NJ, YSA, 1997.

[8] V. Pulkki, M. Karjalainen, and J. Huopaniemi, "Analyzing virtual sound source attributes using a binaural auditory model," *J. Audio Eng. Soc.*, vol. 47, no. 4, pp. 203–217, April 1999.

[9] V. Pulkki and M. Karjalainen, "Localization of amplitude-panned virtual sources I: Sterephonic panning," submitted for publication, 2001.

[10] V. Pulkki, "Localization of amplitude-panned virtual sources II: three-dimensional panning," submitted for publication, 2001.

[11] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," Tech. Rep. #280, MIT Media Lab Perceptual Computing, MA, USA, 1994.

[12] G. Theile and G. Plenge, "Localization of lateral phantom sources," *J. Audio Eng. Soc.*, vol. 25, no. 4, pp. 196–200, April 1977.