Aalto University

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Pulkki, Ville; Pontynen, Henri; Santala, Olli

Spatial Perception of Sound Source Distribution in the Median Plane

Published in:
Journal of the Audio Engineering Society

DOI:
10.17743/jaes.2019.0033

Published: 01/11/2019

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Pulkki, V., Pontynen, H., & Santala, O. (2019). Spatial Perception of Sound Source Distribution in the Median Plane. *Journal of the Audio Engineering Society*, *67*(11), 855-870. https://doi.org/10.17743/jaes.2019.0033

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

OPEN ACCESS Freely available online

V. Pulkki, H. Pöntynen and O. Santala, "Spatial Perception of Sound Source Distribution in the Median Plane"
J. Audio Eng. Soc., vol. 67, no. 11, pp. 855–870, (2019 November.).
DOI: https://doi.org/10.17743/jaes.2019.0033

Spatial Perception of Sound Source Distribution in the Median Plane

VILLE PULKKI, *AES Fellow*, **HENRI PÖNTYNEN, AND OLLI SANTALA**
(Ville.Pulkki@aalto.fi) (Henri.Pontynen@aalto.fi) (Olli.Santala@helimaki.fi)

Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

Spatial perception of concurrently active sound sources was investigated in an exploratory listening experiment. Incoherent noise source distributions of varying spatial characteristics were presented from loudspeaker arrays in anechoic conditions. The arrays were coinciding with the ±45° angular sectors in the frontal median and horizontal planes. The task of the immobile subjects was to report the directions of loudspeakers they perceived emitting sound. The results from median plane distributions suggest that two concurrent sources located along the vertical midline can be perceived individually without resorting to head movements when they are separated in elevation by 60° or more. With source pairs separated by less than 60°, and with more complex physical distributions, the distributions were perceived inaccurately, biased, and spatially compressed but nevertheless not as point-like auditory images.

1 INTRODUCTION

Modern spatial audio reproduction techniques of headphones or loudspeakers seek to control the perceived spatial image as accurately as possible in three dimensions [1]. The techniques utilize either real loudspeakers or virtual loudspeakers implemented through HRTF-processing around the listener in a 3-dimensional setup. Optimally, the techniques should be able to produce perceptions of sound sources with any angular distribution in azimuth and elevation. Thus, understanding the perceptual mechanisms of the listeners is important. Currently, the mechanisms of spatial distribution perception have been studied mainly in the horizontal plane, and this article seeks to shed some light on the corresponding phenomena in the median plane.

A task for spatial audio reproduction techniques is to produce the impression of diffuse reverberant sound fields. In this case listeners should not perceive the spatial distribution of sound concentrated in the directions of the loudspeakers in the setup; instead, an even perception of sound is desired in three dimensions. The requirements from the loudspeaker setup have been studied with systems containing loudspeakers positioned only on the horizontal plane. In such setups, an angular separation of 60° (i.e., six loudspeakers around the listener) has been found to be sufficient to produce a perception of a surrounding diffuse sound field [2]. Corresponding parameter values for loudspeaker arrays positioned on the median plane are currently not known, and a partial aim of the present work is to address this issue.

A more general question of interest in this work is how accurately arbitrary physical distributions of incoherent broad-band sources are perceived when the distributions lie in the median plane. The existing literature [3] suggests very limited or no ability to perceive spatial distribution in such cases. However, the studies have been conducted with relatively short noise bursts, and longer stimuli may allow more accurate spatial perception of sound source distribution.

This article presents an exploratory subjective listening test conducted with a 3-dimensional loudspeaker setup in an anechoic chamber. Separate horizontal and vertical loudspeaker arrays were utilized to create a number of arbitrarily chosen distributions, which were used to probe the human ability to perceive the spatial distribution of sound sources in the frontal horizontal and median planes.

2 BACKGROUND

2.1 Perception of Multiple Concurrently Active Sound Sources

The main signal characteristics that the human auditory system uses to localize sound sources are interaural level and time differences (binaural cues) and spectral features of the sound pressure signal at the eardrum (monaural cues) [4]. Whereas monaural cues are unique for all source directions due to the asymmetric acoustic filtering of incoming sound waves by the head, torso, and pinnae of the

listener, interaural differences are approximately equal for all source directions with rotational symmetry about the interaural axis. Therefore, rather than specifying an explicit source direction, binaural cues define an ambiguous locus of possible source directions that has to be resolved with the aid of monaural cues or head movements. In the case of sound sources located along the vertical midline of the observer, this locus coincides with the median plane as all source elevations along the midline are approximately equidistant from the two ears. The lack of useful directional cues furnished by interaural differences manifests as poorer performance in various auditory tasks as compared to corresponding tasks with horizontally separated sources [5, 3, 6, 7].

The majority of past psychophysical studies assessing auditory perception of concurrent sources have been limited to scenarios involving azimuthally separated sources. The spatial extent of auditory images formed by concurrent horizontally separated sources is affected by the degree of coherence between the sources such that coherent sources form point-like auditory images at a direction within the span of the sources, while incoherent sources are perceived as having a spatial extent that is affected by the degree of intersource coherence [4, 8]. These variations in spatial characteristics have been associated with rapidly fluctuating binaural cues [4, 9]. Segregation of horizontally separated sources has been shown to be facilitated by spectral and temporal dissimilarity, spatial separation, and proximity of the sources to the subject's vertical midline [5, 3, 10–13]; it has been shown that even sources with full spectral overlap can be segregated into a perception of two distinct directions as the azimuthal separation between the sources is increased [3, 13].

As source distributions in the median plane yield negligible interaural differences, any spatial effects related to the vertical extent of auditory images are likely to be related to monaural rather than binaural signal characteristics. In the presence of multiple incoherent sources, the monaural spectrum represents the superposition of the pinna-filtered spectra of the individual sources. This is likely to have a detrimental effect on localization performance if the spectral contours serving as spatial cues are significantly distorted by the superposition. Hypothetically, distortions of the spectral contours salient to localization may become less severe at large source separation angles. Consequently, the directional cues corresponding to the individual sources may be better preserved in the monaural spectrum, allowing for some degree of directional segregation based on monaural cues alone.

In one of the few experiments assessing auditory perception of concurrent median plane sources, Best et al. presented their subjects with two concurrent 150-ms bursts of incoherent white noise in virtual auditory space implemented with individualized binaural filters [3]. The task of the subjects was to report whether they perceived sound from one or two directions. The results showed that subjects were unable to segregate median plane source pairs even at angular separations as large as 90° and generally reported perception of only one source direction [3] (see also [14] for results from a similar free field task). Although a signal duration of 150 ms should allow for the localization of single sources via monaural cues [15], it is in principle possible that a longer signal duration is required for resolving two median plane source directions.

The perception of multiple coherent sources in the median plane has been studied in the context of amplitude panning. In amplitude panning, a signal is fed to loudspeakers with different gains targeting perception of a point-like auditory image between the sources. In [16] it is shown that, when the sources are in the median plane, a virtual source can be formed between the loudspeakers with certain gain ratios; however, relatively often, the perceived direction coincides with one of the loudspeakers.

2.2 Motivation of the Perceptual Test

In the current work we were interested in studying the perception of source distribution driven by monaural cues. We addressed the question of when several sources emit spectrally overlapping incoherent broadband sound in the median plane, where can we expect the virtual sources to appear? In principle, any perceived direction is possible, as the source signals are superimposed in the ears with unknown interference patterns. However, it was assumed that either a spatially distributed source is perceived somewhere in the actual source distribution, or, in the best case, the perceived distribution of sound corresponds to the actual distribution of sources. The listeners reported the loudspeakers that they thought the sound was coming from by using a graphical user interface; the distribution of the responses was then assumed to be indicative of the perceived distribution. The actual distributions were designed to test the accuracy of perception of complex distributions and to probe the magnitude of vertical loudspeaker separation angles suitable for the spatial audio application of reproducing surrounding, reverberant sound fields to a perceptually relevant degree of accuracy.

The test was repeated with equal source-distribution-cases in both horizontal and vertical planes. This enabled the comparison of the results to characterize the added information provided by binaural cues. The results from an earlier study conducted in our laboratory [13] could not be used for this, since many of the test distributions could not be transferred to the median plane in the laboratory available. Additionally, in the earlier tests, head movements were allowed, whereas here they were restricted.

3 METHODS

3.1 Subjects

Twelve staff members from the Department of Signal Processing and Acoustics at the Aalto University School of Electrical Engineering participated in the experiment. The subjects were between 24 and 39 years of age, reported normal hearing, and had extensive previous experience in spatial hearing experiments. The authors did not participate in the experiment. All 12 subjects participated in both experimental sessions.

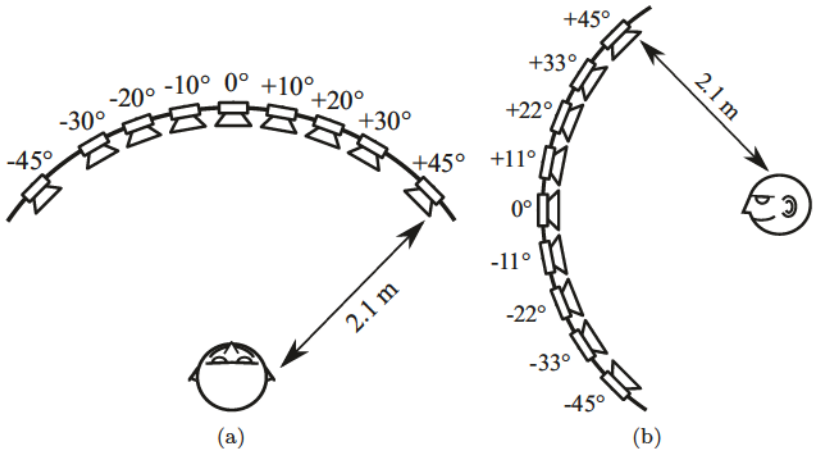


Fig. 1. Loudspeaker setups in the horizontal (a) and vertical (b) tests.

3.2 Experimental Setup

The experiment was conducted in an anechoic chamber equipped with two arched loudspeaker arrays of 2.1-m radius (see Fig. 1) extending along the horizontal and vertical dimensions. The perpendicular arrays intersected at the middle loudspeaker common to both arrays. The listening position was in the middle of the test chamber where the arrays coincided with the azimuthal and median planes of the seated subjects. The horizontal array consisted of nine loudspeakers separated by 10° and 15° intervals as measured from the listening position. The loudspeakers were clearly labeled with integer numbers (–4, ... 4) in both vertical and horizontal arrays, and the loudspeaker directly in front was labeled with number zero. The two endmost loudspeakers at either extreme of the horizontal array were separated by 15° from the rest of the array while the separation between the remaining adjacent loudspeakers was 10°. The vertical array was constructed similarly with nine loudspeakers separated by 11° and 12° intervals. The total angular extent of both arrays was 90° in their respective dimensions. All loudspeakers were individually calibrated to produce the same sound pressure level at the listening position with ±0.5-dB tolerance. Subjects reported their responses to each trial with a tablet computer.

3.3 Stimuli

The stimuli comprised of 44 different source distributions formed by presenting incoherent pink noise from various loudspeaker combinations in either array. Nineteen distributions were formed symmetrically about the subject's midline, and 16 distributions were laterally displaced versions of selected symmetric distributions. In addition, nine point-like stimuli were presented from the individual loudspeakers for control purposes. These 44 source distributions were presented with both horizontal and vertical loudspeaker arrays, totaling 88 different test cases. The total duration of each stimulus was 15 s. Envelopes were controlled by applying 100-ms linear ramps at stimulus onset and offset. The sound pressure level variations between source distributions with different numbers of active sources were minimized by scaling each loudspeaker signal by a factor of $1/\sqrt{n}$, where n denotes the number of sources in the distribution. The A-weighted sound pressure level produced at the listening position by each stimulus was 60 dBA.

3.4 Test Procedure

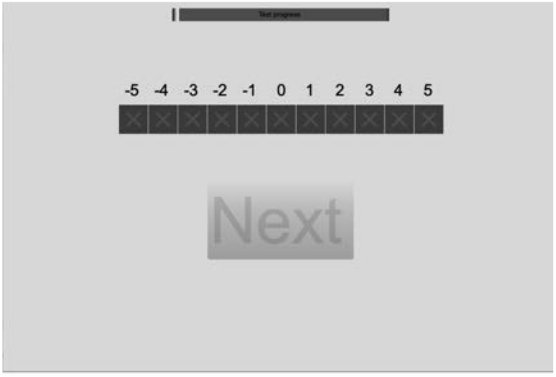
The experiment was conducted in two separate experimental sessions. Half of the subjects assessed azimuthal distributions in the first session and median plane distributions in the second session; this order was reversed for the remaining subjects. Subjects familiarized themselves with the psychophysical task and the response interface in training sessions before both parts of the experiment. All subjects completed 3 trials of all 44 stimuli presented in random order in both experimental sessions.

During the experiment, subjects were seated at the listening position facing the middle loudspeaker common to both arrays. Head movements were restricted by a headrest that ensured correct orientation during the experiment. In each trial the task of the subjects was to report the perceived directions of auditory activity via a tablet computer. The user interfaces are shown in Fig. 2. Each interface consisted of 11 toggle buttons corresponding to the nine angular directions of the loudspeakers in the active array and two additional directions, marked in the interface with labels −5 or 5, that corresponded to directions beyond either extreme of the array.

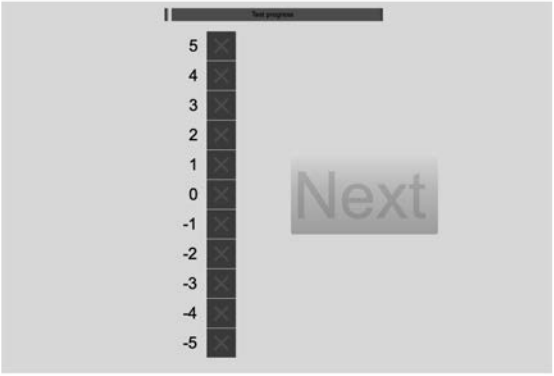
The available response options encompassed all combinations of the 11 directions on the response interface that included at least one active direction. Subjects could report their responses before the termination of each 15 s stimulus; early responses truncated the ongoing stimulus with a 100 ms linear ramp. The average duration of the experimental sessions was approximately 30 minutes.

3.5 Visual Representation of the Results

The results are presented as barplots of the pooled responses and heat maps of co-occurrence matrices $\mathbf{C(n, m)}$, where n and m indexes denote the directions (−X, −45°, . . . , +45°, +X) corresponding to the responses available to the subjects, where (±X) denotes the directions beyond



(a) Horizontal test.



(b) Vertical tests.

Fig. 2. The graphical user interfaces for listening tests.

the extremes of the arrays. The barplot presentation shows how often each loudspeaker has been marked to produce sound. However, in cases where the responses are distributed over several directions, the barplot presentation gives quite coarse indication about the distribution. The perceived distribution might be even or composed of several centers of gravity in different directions. A somewhat better view of the perception of the subjects is obtained with heat maps, in which an even perceived distribution is shown as a highlighted square, whereas the perception of several centers of gravity are shown as highlighted spots on the map.

The co-occurence matrices for each distribution type were computed by combining the binary response vectors to all 36 trials (pooled across subjects) into a response matrix $\mathbf{R}(i, n)$, where the rows denoted with i correspond to the response vectors from individual trials (1–36) and the columns, denoted with n, correspond to the 11 possible response directions.

The entries of the co-occurence matrices were computed from the response matrices of each distribution type with the formula:

$$\mathbf{C}(n, m) = \sum_{i=1}^{36} \mathbf{R}(i, n)\mathbf{R}(i, m), \tag{1}$$

where $n, m \in \{1, \ldots, 11\}$. This yields an 11×11 symmetric matrix with both rows and columns corresponding to response directions.

The higher the value of $\mathbf{C}(n, m)$ with a specific (n, m) pair, the more often they have been indicated to produce sound at the same time. Off-diagonal entries in co-occurrence matrices correspond to the number of responses containing both of the two directions corresponding to the row and column number of the entry. Diagonal entries correspond to the total number of responses recorded for individual directions. In the heat map representations (e.g., Fig. 3), the black crosses along the bottom-left to top-right running diagonals of the matrices correspond to the spatial arrangement of the loudspeakers comprising each stimulus. The cells along this diagonal also correspond to the barplots of pooled responses attached to the side or top of the matrices, i.e., they represent the total number of times a direction was present in the responses. The maximum value of the barplots and heat map cells is 36, corresponding to 12 subjects completing 3 trials of each stimulus. Additionally, the mean number of reported directions per trial and the standard deviation across subjects are presented for each distribution type accompanying the corresponding co-occurrence matrix.

4 RESULTS

As the number of total distributions tested was relatively high (88 in total), the inclusion of all the of data was not deemed necessary for this article. The results were analyzed, and only the results relevant to the conclusions of the work are presented here. For interested readers, all of the plots can be found in supplementary material in the Appendix.

4.1 Horizontal Plane Array of Loudspeakers: Results

Six response distributions resulting from the azimuthal tests are shown in Fig. 3. While in the previous tests [13] the subjects were allowed to move their heads, in this case they were immobilized. This could have potentially led to lower accuracy in the perception of spatial distribution as dynamic cues were not available, but nevertheless, the present results are qualitatively similar to those presented previously in [13]. Azimuthally broad and evenly distributed sources were perceived to be wide, as shown in Figs. 3(a) and 3(b). The response distribution edges are biased slightly towards the median plane, as was the case in the previous study, although the previous results showed a clear emphasis on the edges of the distribution, while in the present results, the distribution of the responses seem relatively flat.

Two sources separated by 90° in azimuth clearly resulted in perception of two separate directions, as shown in Fig. 3(c). When three sources are separated by 45° or 30° angles to the adjacent source(s), as in Figs. 3(d) and 3(e), it can be seen that the perception of individual sources is less clear, and the gaps in the response distribution become less prominent. The present results cannot be directly compared with the results in [13], since such distributions were not tested in that study.

The case shown in Fig. 3(f) is included as an example of perception of a source with complex spatial distribution.

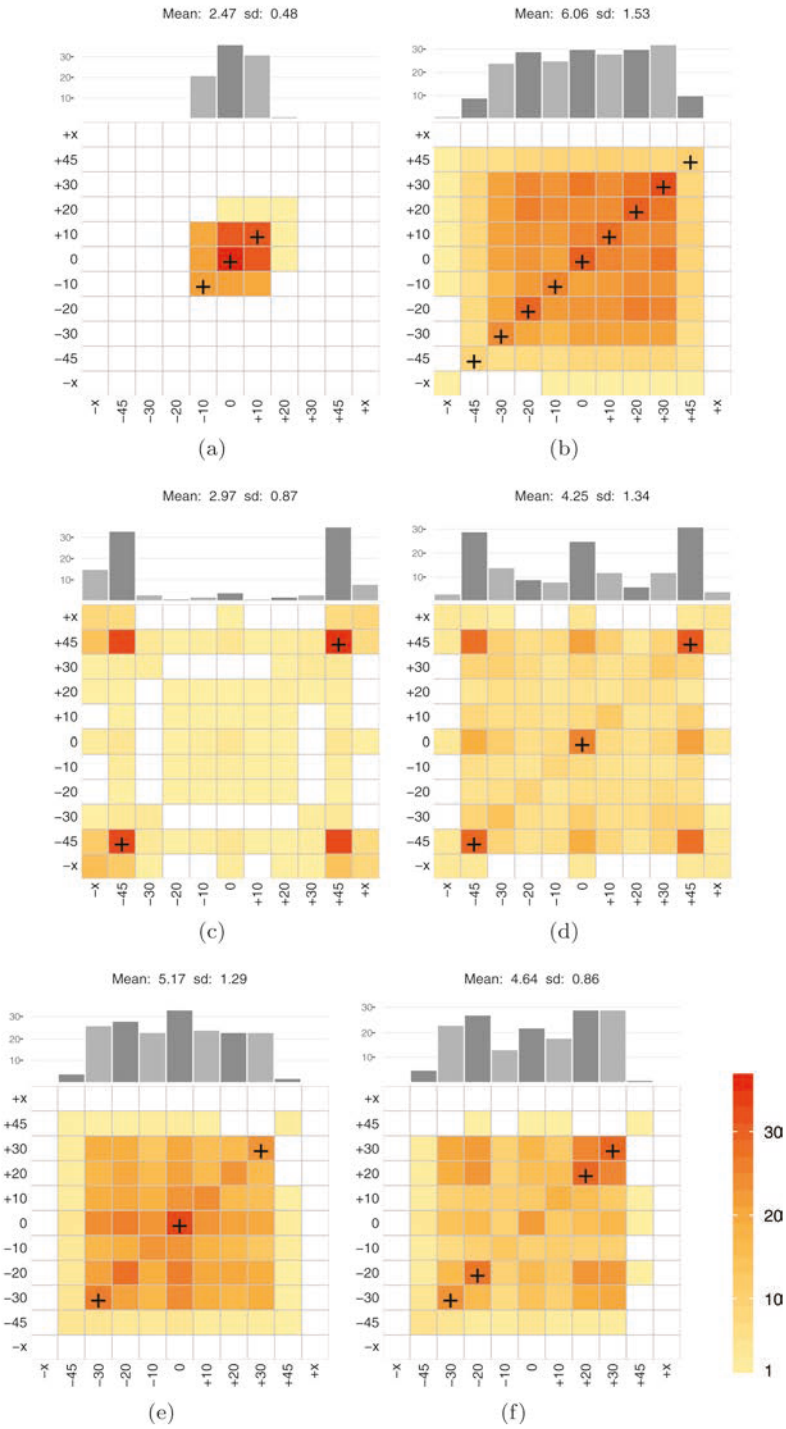


Fig. 3. Results from the experiments in the horizontal plane. The listeners indicate which loudspeakers are emitting sound from a loudspeaker array. The crosses indicate the directions of concurrent loudspeakers emitting incoherent pink noise. The histogram shows the distribution of the responses, where the two shades of gray are to differentiate visually the bars from each other. The colors of the cells of the heat maps denote how often both of a loudspeaker pair were indicated to emit sound. Mean and sd denote the average number of marked directions and standard deviation across the 12 subjects. Forty-four different source distributions in the horizontal plane were used as test cases. The results from only selected cases are shown in this article and the rest are shown in the Appendix.

It can be seen that the width of the indicated distribution matches with the actual distribution, while the fine details of the source distribution are reflected only vaguely in the response distribution.

It can be seen in the data shown here and in the supplementary material in the Appendix that the average number of loudspeakers indicated to emanate sound matches only coarsely with the number of actual sources. In addition, with higher number of loudspeakers a negative bias is seen in the responses. For example, the maximum average number of indicated loudspeakers was approximately six, although the maximum number of loudspeakers was nine.

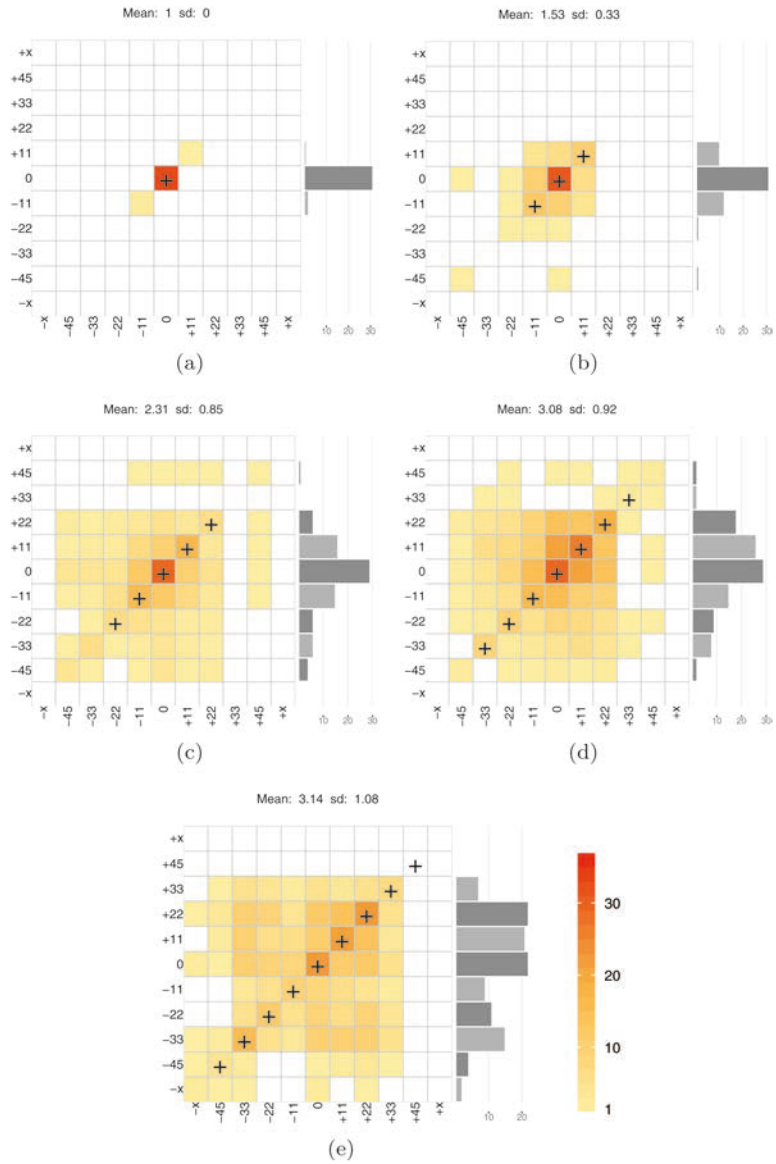


Fig. 4. Results from the experiments in the median plane. The results from tests with increasing vertical extent of evenly distributed sources in the median plane. Forty-four different source distributions in the horizontal plane were used as test cases. The results from only selected cases are shown in this article, and the rest are shown in the Appendix.

4.2 Median Plane Array of Loudspeakers: Perceived Distribution of Sound Source

In the case of the median plane results, the first question discussed is if a vertically extended source distribution causes perception of a continuous, vertically extended but horizontally concise sound source. Fig. 4 shows the results from source distributions with varying amounts of adjacent active sources. The change is clear in the heat maps from stimuli consisting of a single active loudspeaker to stimuli consisting of three and five active loudspeakers, where the subjects have apparently perceived a more distributed source when the number of sources increased. However, even with five sources the responses are concentrated near the center of the physical distribution. When the distribution is made broader to seven and nine sources, the changes in the results can be said to be minor. Moreover, the responses concentrate on directions at 0° and above rather than below the horizontal plane, even though the source distributions were symmetrical around 0°. The prominent feature in the response distributions is that independent of the actual number of active sources, the heat maps in Figs. 4(c), 4(d), and 4(e) show that the listeners have favored only three of the response directions prominently.

The results thus suggest that sources distributed along the median plane are perceived to be more extended vertically than point-like sources; however, the response distribution is compressed considerably as compared to the actual source distribution. This is also reflected in the average number of loudspeakers indicated to emanate sound, being only 3.14 ± 1.08 with nine-loudspeaker case. Furthermore, as expected, the accuracy of spatial perception is prominently lower with vertical distributions than with the corresponding azimuthal distributions shown in Figs. 3(a) and 3(b).

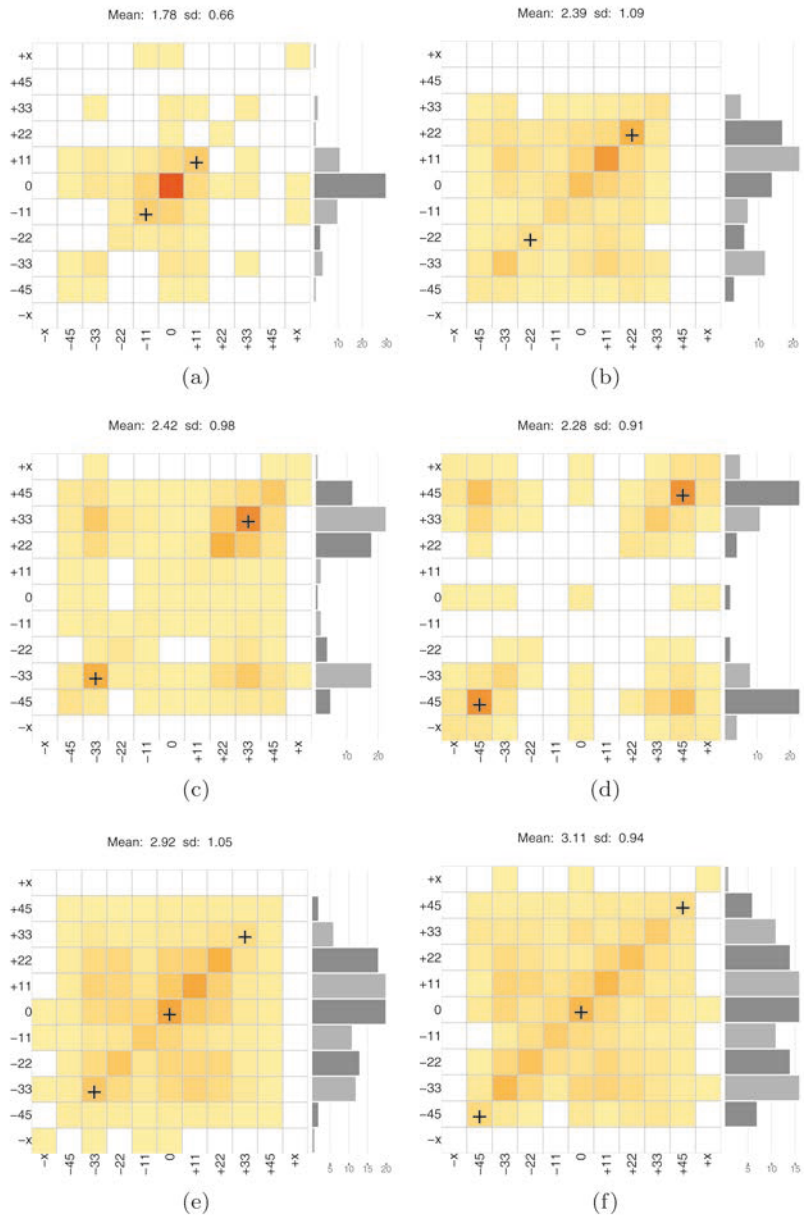


Fig. 5. As in Fig. 4, but with tests with two or three spaced sources.

The present work was motivated by the spatial audio application of reproducing reverberant sound fields to a perceptually relevant degree of accuracy by using loudspeakers spaced sparsely in elevation. In this application, the goal is to avoid situations in which listeners are able to perceive the directions of the loudspeakers rather than perceiving a diffuse auditory image that can encompass also the gaps in the sparse array. Fig. 5 shows the results with two and three sources with different spacings. Figs. 5(a) to 5(d) show the results with two sources separated symmetrically around the horizontal plane. It can be seen that while with 22° separation the perceived distribution is clearly unimodal, with 66° and 90° separations the heat maps show that the listeners have indicated the correct source pairs. This is pronounced in the case with 90° separation, as the sources around the midline have not been indicated to produce sound. When a third source is added to the array, as in Figs. 5(e) and 5(f), the subjects have not detected the presence of any of the sources clearly, and consequently, the response distributions are relatively flat.

Fig. 6 shows the results with arbitrarily chosen complex distributions. Most of the results show that the perceived distribution is relatively far from the actual distribution, and biasing and/or compression effects can be seen in the results. Some of the distributions contained a gap in the physical distribution, which is perceived somewhat consistently only in Fig. 6(b). When these results are compared with corresponding horizontal results plotted in Figs. 3(e) and 3(f), it can be seen that the horizontal results have better correspondence to the actual edge-to-edge span of the source than the results from the median plane, which are often biased to the elevated end of the distribution.

The results shown in these figures suggest that the human ability to accurately hear the spatial properties of complex source distributions in the median plane is limited to cases in which two compact sources are

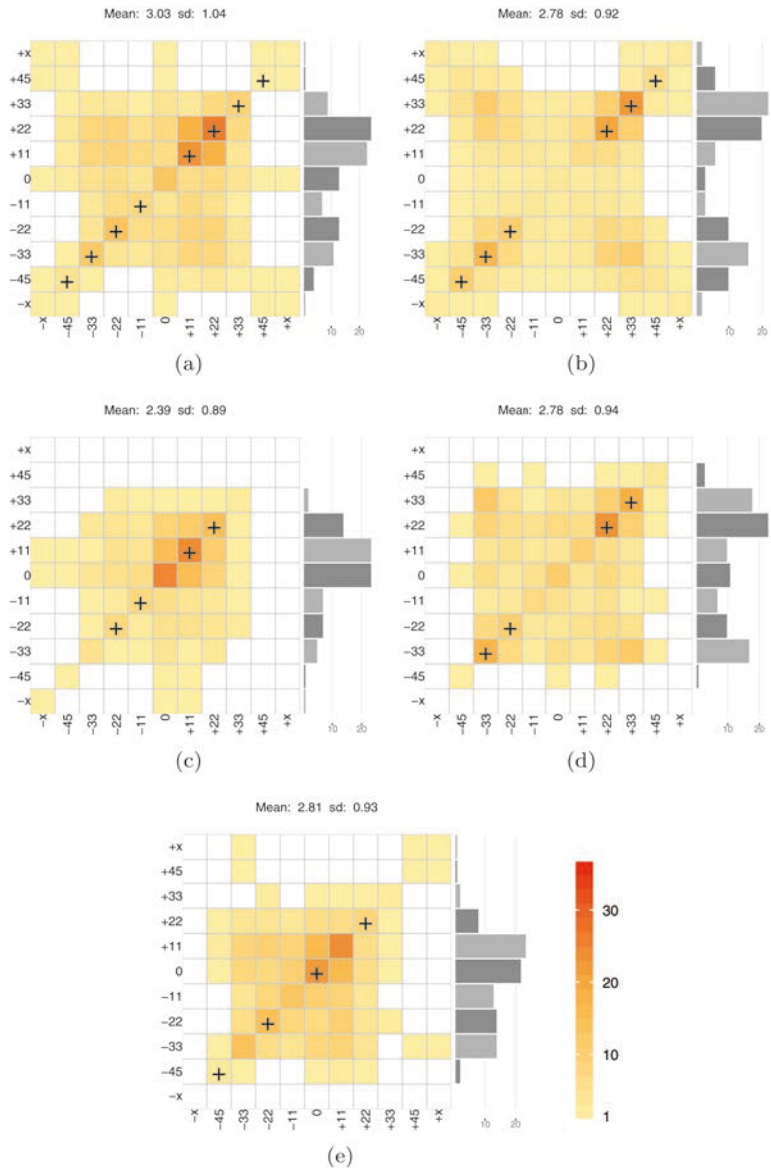


Fig. 6. As in Fig. 4, but with tests with complex distributions.

separated by at least 60°. This also suggests that the separation in elevation angle between adjacent sources in 3-dimensional loudspeaker setups should be clearly less than 60° in the median plane, as in that case, listeners should not be able to perceive the directions of individual loudspeakers.

4.3 Modality of Response Distributions

The results in Figs. 5(c) and 5(d) show bimodal response distributions and co-occurrence matrices with distinct response-pair clusters. This suggests that angular separations above approximately 60° between loudspeakers were apparently large enough to enable consistent segregation of median plane source pairs. In this regard, the current results are in contrast with those from the study by Best et al. [3]. Whereas the conclusion reached in [3] was that subjects could not segregate median plane sources even with angular separations as large as 90°, the results from the current study suggest that segregation is possible already with the separation value of 66°. A possible reason for this discrepancy is the difference in stimulus type and stimulus durations between the experiments. While in the current study subjects could be exposed to individual stimuli for as long as 15 s, the noise bursts used by Best et al. were limited to 150 ms. Although the stimulus length should allow for the localization of single sources via monaural cues [15], it is possible that a longer time frame is needed for extracting stable monaural cues from the monaural spectrum yielded by multiple incoherent sources. Further, whereas the study by Best et al. [3] employed headphone stimuli presented through individualized HRTFs, the current study used natural stimuli presented over loudspeakers. While the use of loudspeakers on the one hand guaranteed that natural spatial cues were available for all subjects, it also reduced the amount of experimental control we had over binaural differences. Therefore, it is unclear whether the discrepancy between the current data and that presented in [3] is a result of segregation being facilitated by the loudspeaker

presentation of longer stimuli or the presence of small binaural differences that were not controlled for in the present study. In any case, the current data show that the subjects were able to segregate two point-like sources under natural listening conditions in a much more consistent manner than data from previous experiments has shown.

5 CONCLUSIONS

Spatial perception of source distributions on the frontal median and horizontal hemiplanes by immobile listeners was investigated in an exploratory two-part listening experiment with horizontal and vertical loudspeaker arrays in anechoic conditions. The conclusions of the work are as follows:

1. The perception of sources with complex spatial distribution in the horizontal plane is not degraded prominently when head movements are restricted.
2. Two concurrent sources emitting incoherent noise in the median plane can be perceived individually if the sources are separated in elevation by approximately 60° or more.
3. Median plane source pairs separated in elevation by less than 60° tend to form auditory images at intermediate directions between the sources.
4. Spatial perception of complex distributions of sources in the median plane departs prominently from the spatial properties of the actual distribution and has notably lower accuracy than the perception of corresponding source distributions in the horizontal plane.

Overall, the results suggest that in 3-dimensional multichannel loudspeaker setups, the separation of loudspeakers should be of the order of 45° in the frontal median plane to produce perception of evenly distributed sound in the median plane. Consequently, if similar accuracy holds in all directions in the median plane, eight evenly spaced loudspeakers are needed in the median plane for 3-dimensional loudspeaker setups in order to avoid the detection of individual loudspeaker directions in spatial audio applications aiming at reproduction of diffuse sound.

6 ACKNOWLEDGMENT

This work has been supported by the Academy of Finland.

7 REFERENCES

[1] V. Pulkki and M. Karjalainen, *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics* (John Wiley & Sons, 2015).

[2] K. Hiyama, S. Komiyama, and K. Hamasaki, "The Minimum Number of Loudspeakers and Its Arrangement for Reproducing the Spatial Impression of Diffuse Sound Field," presented at the *113th Convention of the Audio Engineering Society* (2002 Oct.), convention paper 5696.

[3] V. Best, A. van Schaik, and S. Carlile, "Separation of Concurrent Broadband Sound Sources by Human Listeners," *J. Acoust. Soc. Amer.,* vol. 115, no. 1, pp. 324–336 (2004).

[4] J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization*, 2nd ed. (MIT Press, Cambridge, MA, 1997).

[5] D. R. Perrott and K. Saberi, "Minimum Audible Angle Thresholds for Sources Varying in Both Elevation and Azimuth," *J. Acoust. Soc. Amer.*, vol. 87, no. 4, pp. 1728–1731 (1990).

[6] J. C. Makous and J. C. Middlebrooks, "Two-Dimensional Sound Localization by Human Listeners," *J. Acoust. Soc. Amer.*, vol. 87, no. 5, pp. 2188–2200 (1990).

[7] J. C. Middlebrooks, "Narrow-Band Sound Localization Related to External Ear Acoustics," *J. Acoust. Soc. Amer.*, vol. 92, no. 5, pp. 2607–2624 (1992).

[8] M. B. Gardner, "Image Fusion, Broadening, and Displacement in Sound Location," *J. Acoust. Soc. Amer.*, vol. 46, no. 2B, pp. 339–349 (1969).

[9] M. J. Goupell and W. M. Hartmann, "Interaural Fluctuations and the Detection of Interaural Incoherence. iii. Narrowband Experiments and Binaural Models," *J. Acoust. Soc. Amer.*, vol. 122, no. 2, pp. 1029–1045 (2007).

[10] D. R. Perrott, "Discrimination of the Spatial Distribution of Concurrently Active Sound Sources: Some Experiments with Stereophonic Arrays," *J. Acoust. Soc. Amer.*, vol. 76, no. 6, pp. 1704–1712 (1984).

[11] D. R. Perrott, "Concurrent Minimum Audible Angle: A Re-Examination of the Concept of Auditory Spatial Acuity," *J. Acoustics . Soc. Amer.*, vol. 75, no. 4, pp. 1201–1206 (1984).

[12] P. L. Divenyi and S. K. Oliver, "Resolution of Steady-State Sounds in Simulated Auditory Space," *J. Acoust. Soc. Amer.*, vol. 85, no. 5, pp. 2042–2052 (1989).

[13] O. Santala and V. Pulkki, "Directional Perception of Distributed Sound Sources," *J. Acousti. Soc. Amer.*, vol. 129, no. 3, pp. 1522–1530 (2011).

[14] T. Kim and H. Pöntynen, "Perception of Vertically Separated Sound Sources in the Median Plane," presented at the *2019 AES International Conference on Immersive and Interactive Audio* (2019 Mar.), conference paper 46.

[15] J. Vliegen and A. J. Van Opstal, "The Influence of Duration and Level on Human Sound Localization," *J. Acoust. Soc. Amer.*, vol. 115, no. 4, pp. 1705–1713 (2004).

[16] V. Pulkki, "Localization of Amplitude-Panned Virtual Sources ii: Two-and Three-Dimensional Panning," *J. Audio Eng. Soc.*, vol. 49, pp. 753–767 (2001 Sep.).

APPENDIX

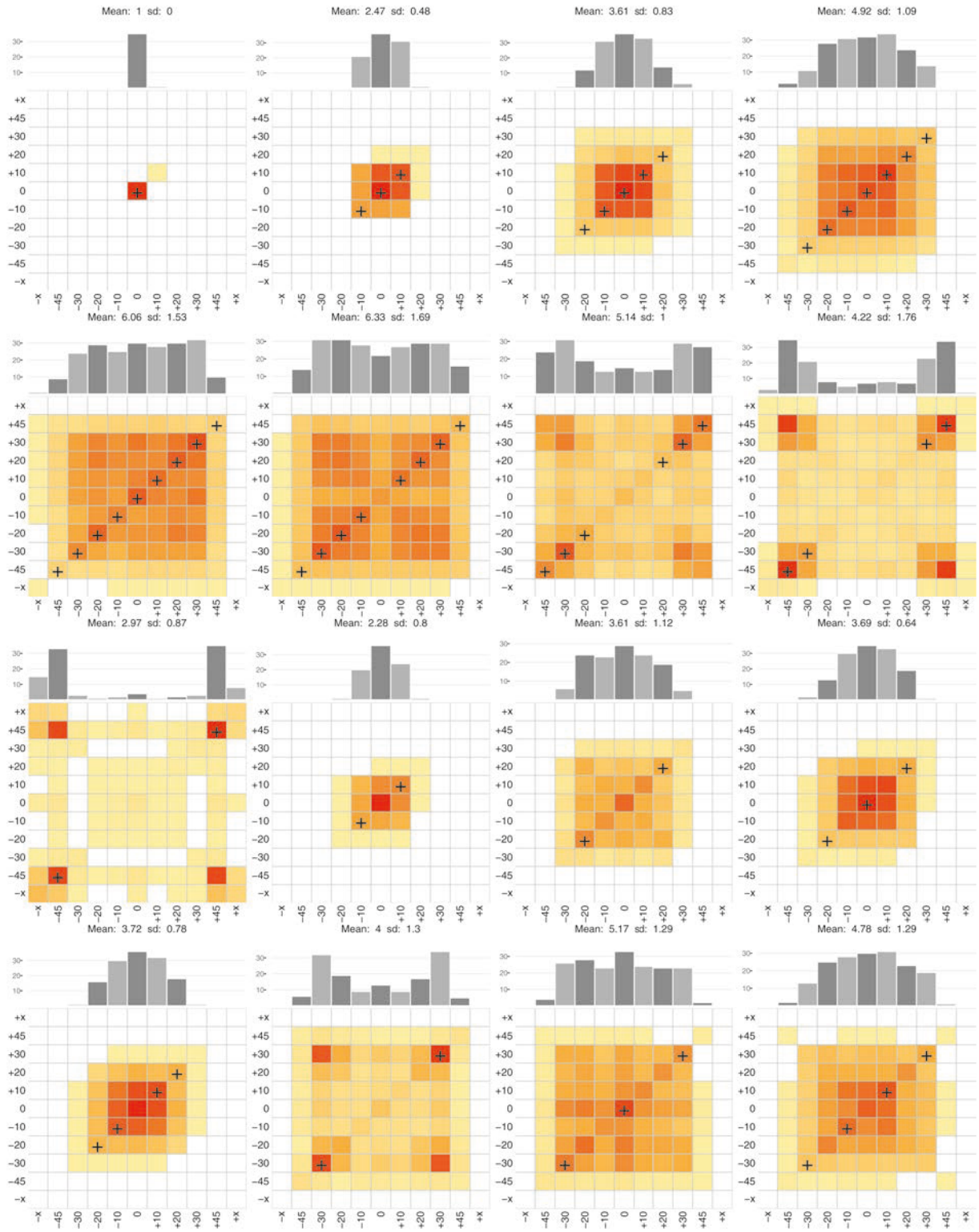


Fig. 7. Horizontal test results 1.

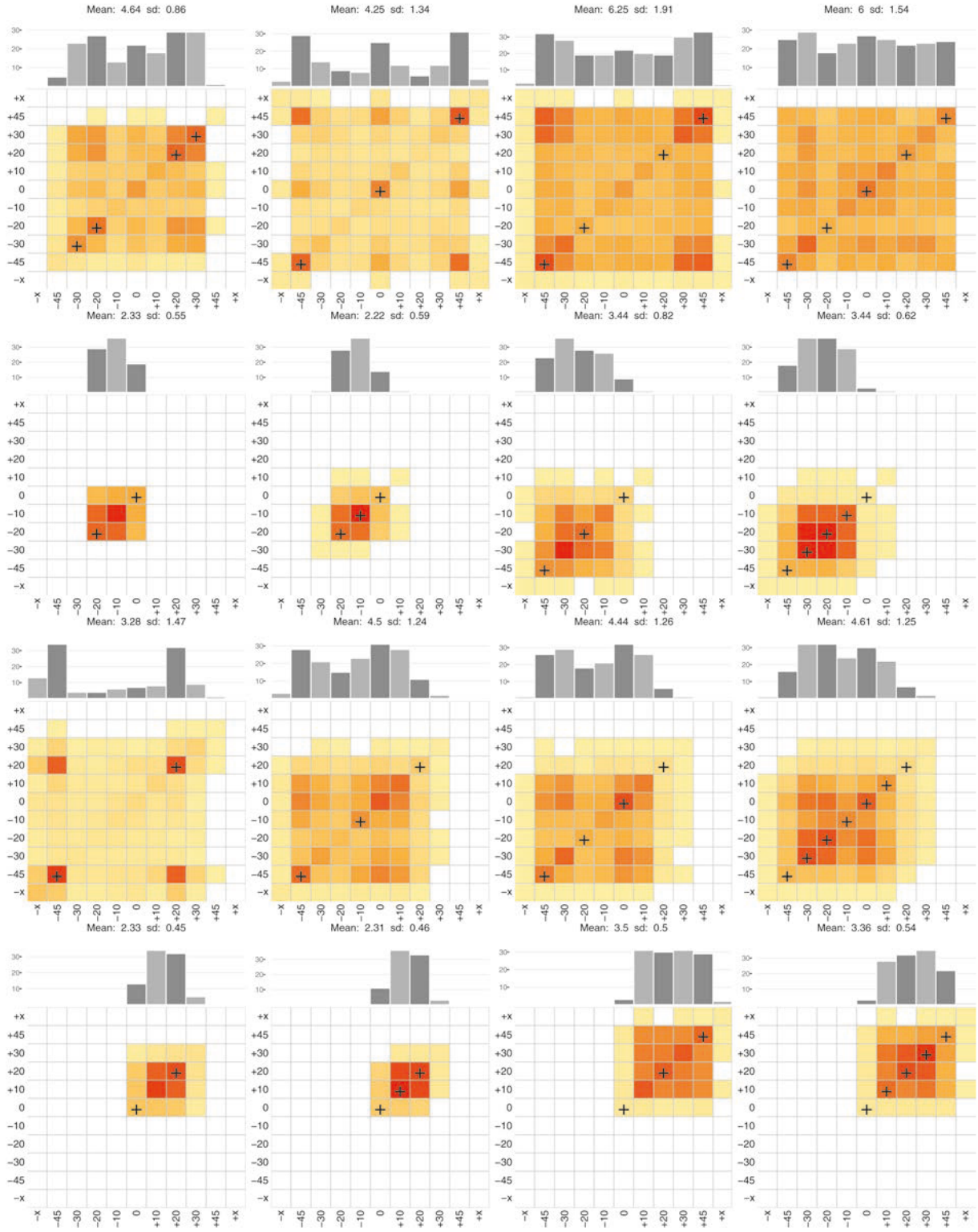


Fig. 8. Horizontal test results 2.

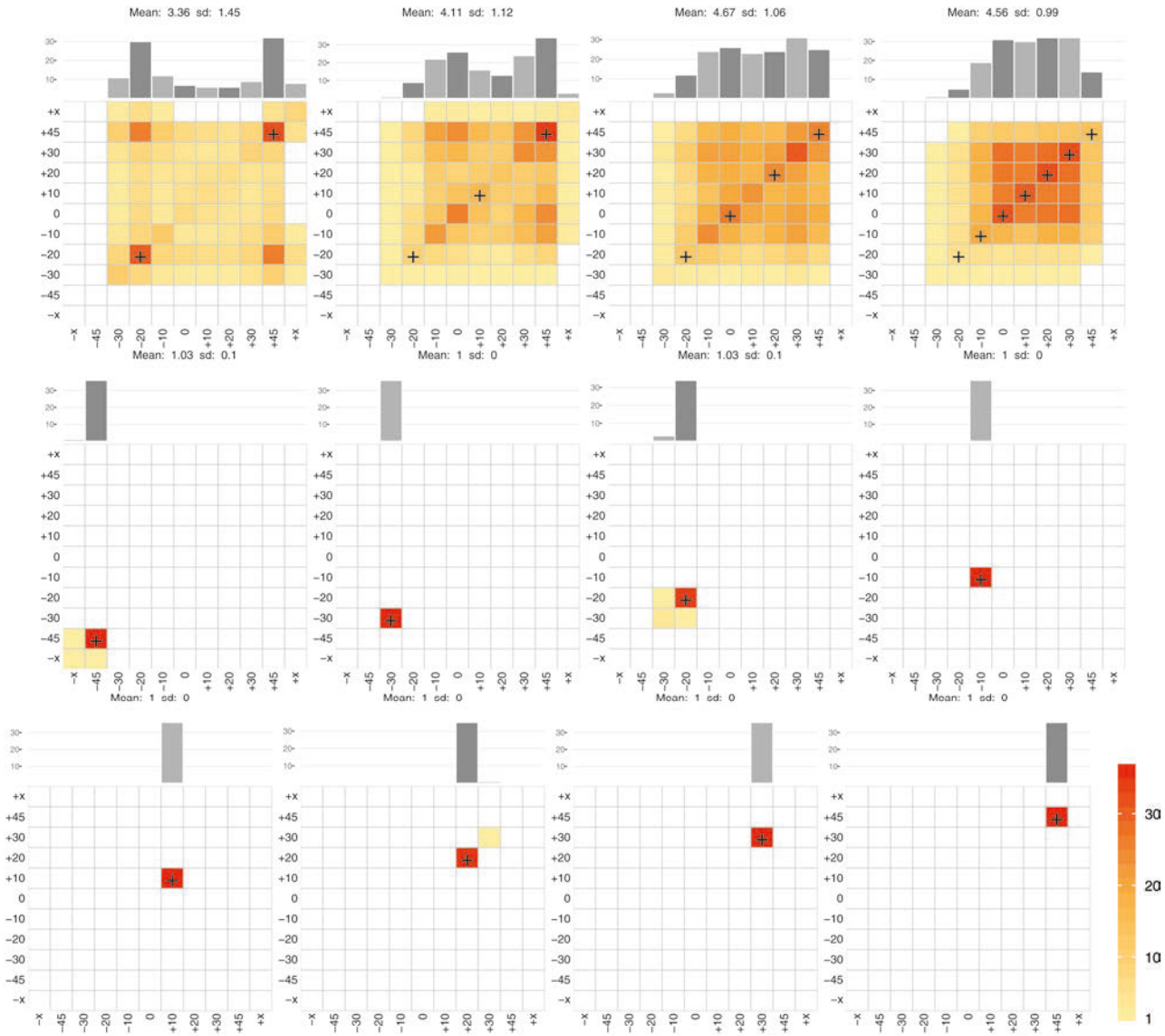


Fig. 9. Horizontal test results 3.

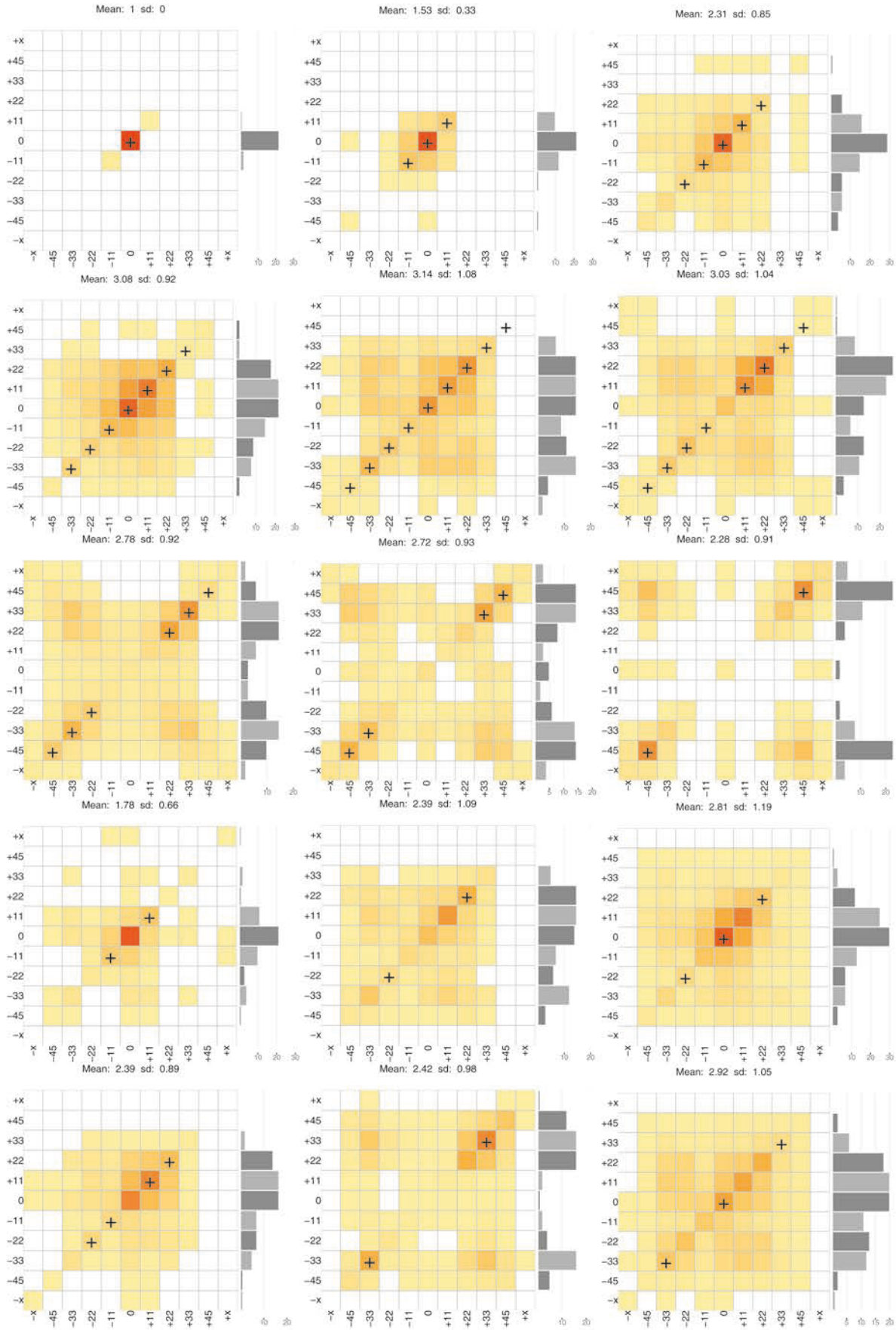


Fig. 10. Vertical test results 1.

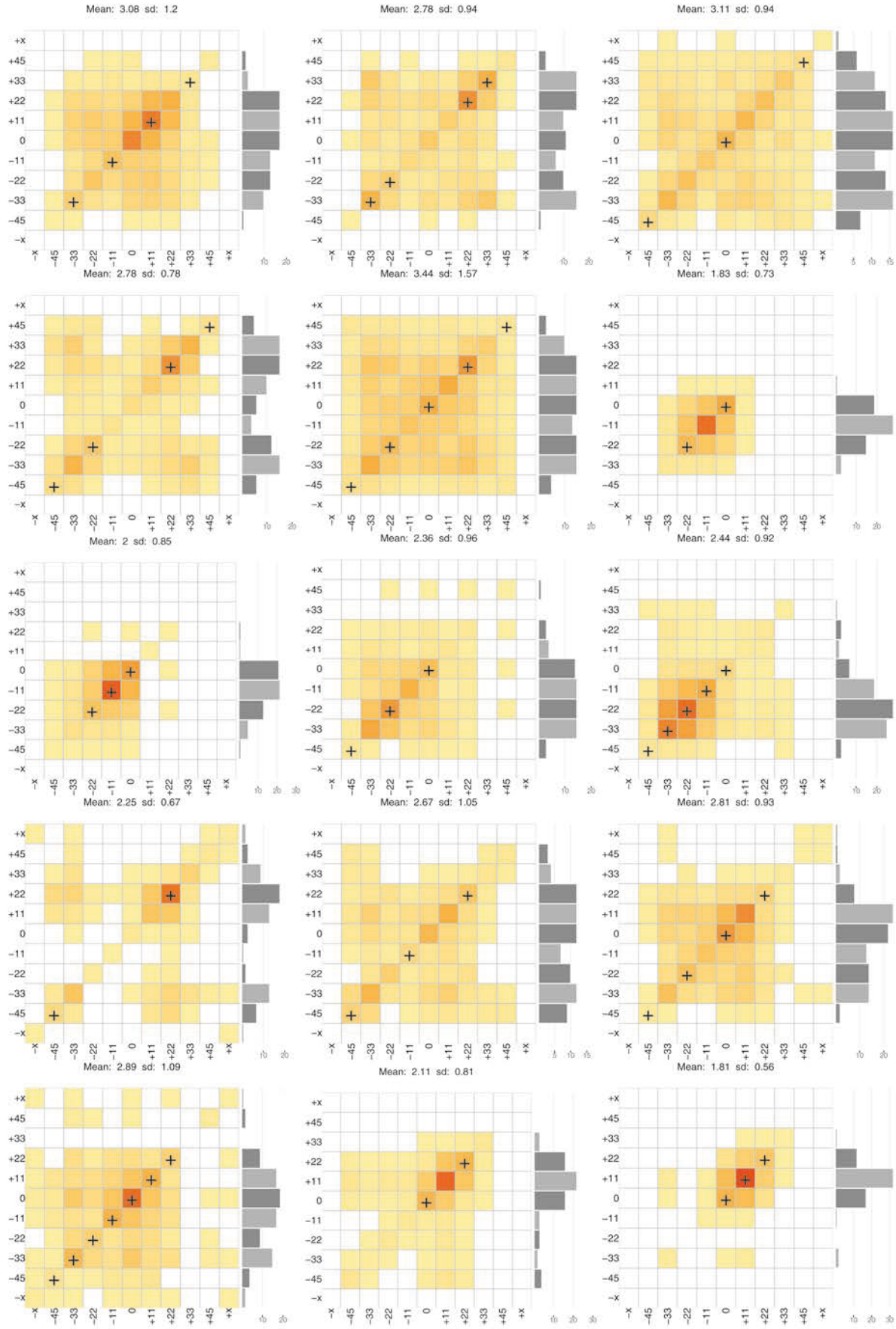


Fig. 11. Vertical test results 2.

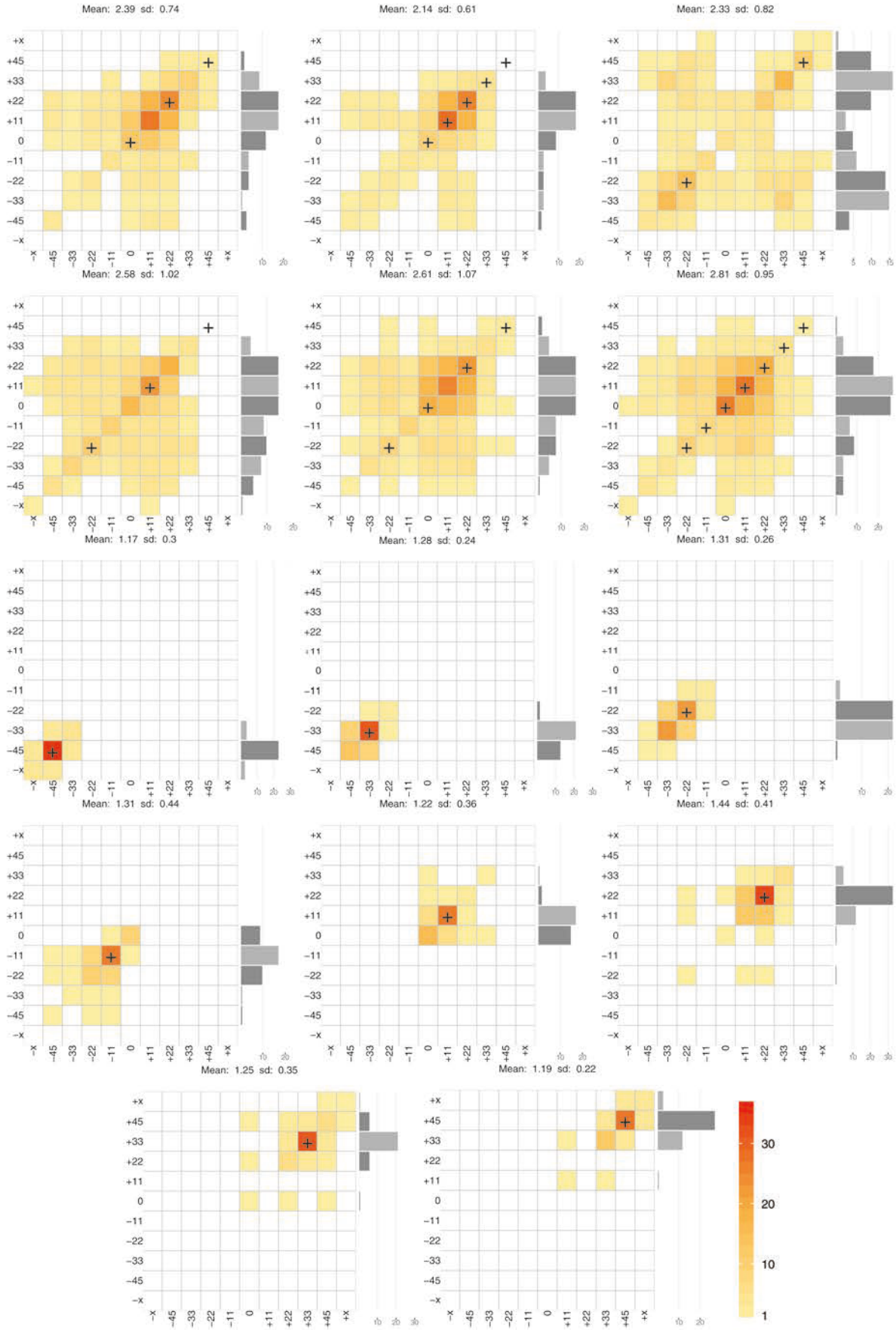


Fig. 12. Vertical test results 3.

THE AUTHORS

Ville Pulkki

Henri Pöntynen

Olli Santala

Ville Pulkki is an associate professor in the Department of Signal Processing and Acoustics in Aalto University, Helsinki, Finland. He has been working in the field of spatial audio for over 20 years. He developed the vector-base amplitude panning (VBAP) method in his PhD (2001), and directional audio coding after the PhD with his research group. He has also contributions in perception of spatial sound, in laser-based measurement of room responses, and in binaural auditory models. He has received Samuel Warner memorial medal from SMPTE and Silver Medal from AES. He enjoys being with his family, building his summer house, and performing in musical ensembles.

•

Henri Pöntynen received his B.Sc. and M.Sc. (Tech.) degrees in electrical engineering from Aalto University in 2014 and 2015 respectively. He is currently pursuing a Ph.D. at the same institution in the field of auditory perception. His research interests include cross-modal effects in spatial hearing, auditory scene analysis and cortical representation of auditory space. During his free time, he enjoys various forms of art, spending time in nature, biking and cooking.

•

Olli Santala received his M.Sc. (Tech) degree in 2009 from the former Helsinki University of Technology, now called Aalto University. This work was continued in the doctoral dissertation that was completed in 2015 and titled Perception and auditory modeling of spatially complex sound scenarios, granting Santala the degree of D.Sc. (Tech). Since the beginning of 2016, he has been pursuing a career at Helimäki Acoustics (a part of Sitowise) as an acoustical consultant specializing on building acoustics. Additionally, he is a member of the national standardization committee on acoustics and the board of the Acoustical society of Finland. In his spare time, he sings in a vocal ensemble with his friends.