Perception of spatial distribution of wide sound sources

Olli Santala and Ville Pulkki

Department of Signal Processing and Acoustics, Helsinki University of Technology (TKK), FI-02015 TKK, Finland

The perception of the spatial distribution of sound sources composed from multiple loudspeakers emitting continuous signal is studied in this article by conducting two listening tests. The tests were performed in an anechoic chamber using 15 loudspeakers evenly distributed in frontal horizontal directions equidistant from the listener. In the first test, various sound source distributions such as sound sources with varying widths and wide sound sources with gaps in the distribution were used to emit uncorrelated pink noise. The subjects were asked to report which loudspeakers emit sound according to their own perception. In the second test, noise signals with different bandwidths as well as sinusoids were used as stimuli. These were presented using loudspeaker combinations with different number of loudspeakers spaced evenly on the frontal horizontal plane. The results of both tests are discussed.

INTRODUCTION

When several loudspeakers are placed close to each other and emit sound simultaneously, it may be hard to distinguish single loudspeakers from the ensemble. Rather, the loudspeakers may form a wide sound source that is perceived as a single sound event. In this article, the perception of spatial distribution of such ensembles is studied in the absence of reflections. A number of previous such studies exist. In a headphone experiment by Perrott and Buell (1981), the perceived spatial width was found to increase as the loudness or duration of the sound was increased. With loudspeakers, Cabrera and Tilley (2003) found the effect of loudness to be similar but smaller than with headphones.

Mason *et al.* (2005) showed that perceived width decreases as frequency increases. Frequencydependency was further studied with loudspeakers by Hirvonen and Pulkki (2006) by using wide band noise that was emitted from different loudspeakers. It was shown that the highest and lowest frequencies of the stimuli affected the spatial perception more than the middle frequencies. In addition, Perrott (1984) observed that discrimination of differences in the spatial distribution of concurrently active sound sources was at its best when the frequency difference of those sound sources was 3%. The task was reliably performed only for frequencies below 1500 Hz, i.e., in the frequencies where interaural time difference (ITD) is said to be dominating the localization (Hirvonen, 2007).

In an article by Hiyama *et al.* (2002) it was tested how similarly sound events are perceived when either 24 loudspeakers or different loudspeaker setups with less loudspeakers are used to produce them. The loudspeakers surrounded the listener at

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

the height of the listeners' ears. With white noise, the results indicated that six evenly distributed loudspeakers produced a very similar perception as 24 loudspeakers. When only four or three loudspeakers were used, the perception was significantly different from the cases where more loudspeakers were used.

EXPERIMENT 1

Experimental setup

Both tests were conducted in an anechoic chamber equipped with a multichannel reproduction system. The test setup, illustrated in Fig. 1, consisted of 15 loudspeakers that were evenly distributed in frontal horizontal directions equidistant from the test subject. The loudspeakers were separated by 15°, thus covering the azimuth sector from -105° to 105°. However, only 13 loudspeakers were actually used in the test to produce sound. The farthest ones on both sides were inactive and were present in order to make it possible to register perceptions equally on any side of the actual sound source. When all 13 loudspeakers emitted sound they formed a physical sound source that was 184° in width. Each loudspeaker is interpreted as a 15° wide element of a spatially distributed sound source, i.e., a gap of one loudspeaker corresponds to a gap of 15° – not 30°.

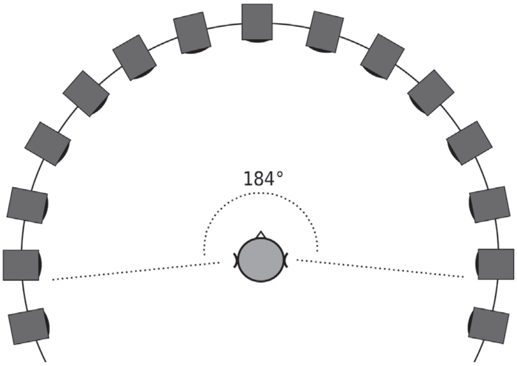

Fig. 1: The loudspeaker setup used in the listening tests. 15 loudspeakers were evenly distributed in frontal horizontal directions equidistant from the listener, forming a 184° wide sound source.

Stimuli

The stimuli were pink noise, and each loudspeaker was driven with independent noise signals so that they were uncorrelated. The total stimulus length was 1000 ms – a 100 ms fade-in, 800 ms constant loudness and a 100 ms fade-out. In each test case, a selected combination of loudspeakers emitted sound at equal loudness level. All the test cases had equal loudness level as well – regardless of the number of loudspeakers emitting sound the overall loudness level was always the same at listener position. The stimulus was constantly repeated until the subject gave his answer, i.e., 1 second stimulus and 1 second silence repeatedly followed one another.

Test design and procedure

21 different loudspeaker combinations were selected as the test cases, and they can be divided into four groups that focus on slightly different details. With these cases, the accuracy of perceiving fine details in spatially distributed sound sources was tested. The test hypothesis was that the resolution of spatial distribution perception is not adequately high to accurately perceive these wide sounds sources and gaps in them.

The first group tested the accuracy of auditory width perception and if it is possible to perceive the sound source as a continuous wide sound event. There were 1 to 13 sound-emitting loudspeakers symmetrically around the center. The second group tested the perception of a hole or a gap in a sound source, or conversely, the resolution of perceiving two separate wide sound sources. Now, a varying number of loudspeakers in the center were not active.

The third and fourth group tested the perception of complex spatially distributed sound sources. The third group included so-called chessboard-type combinations, where every other loudspeaker was emitting sound and every other not emitting. The fourth group consisted of cases where there were two gaps – or conversely, three separate wide sound sources.

The 21 test cases were presented in a randomized order twice for each subject in two separate runs. The subjects were told that in each test case, any of the 15 loudspeakers may emit sound so that any loudspeaker combination is possible. The task was to identify all the loudspeakers that emitted sound according to the subjects' own perception and mark those loudspeakers on a touch screen. Aside from head rotation, no movement was allowed.

Ten voluntary subjects participated in the test. All the subjects were staff or students in the Department of Signal Processing and Acoustics of Helsinki University of Technology and none reported any hearing defects. The authors did not participate in the test.

Results and statistical analysis

A histogram of the results is presented in Fig. 2. The loudspeakers that were emitting sound are marked with a black box, and the gray bars represent the number of times the subjects marked the loudspeakers in question as emitting sound. The y-axis ranges up to 20 as there were ten subjects who each answered twice, thus making a total of 20 responses possible.

The results were analyzed using a statistical analysis method, Kolmogorov-Smirnov (K-S) goodness-of-fit test (Massey, 1951), in order to determine which cases have statistically different distributions and which have statistically similar distributions. In the K-S test, two case distributions at a time are compared. In short, the test gives a *p*-value, and the higher it is, the more similar the two compared distributions are. First, a weighting was done to all the subjects' answers in all cases so that each marked loudspeaker received a weighting coefficient depending on the total number of marked loudspeakers in that particular answer. After that, the weighted histogram distributions of all the cases were compared to each other with the K-S test.

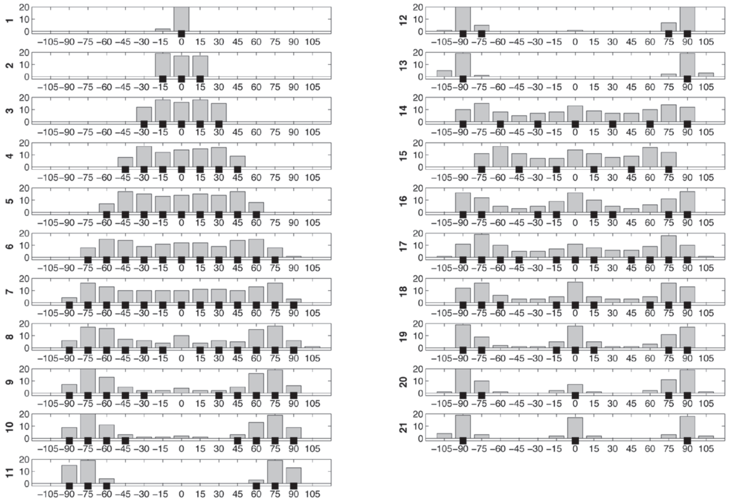

Fig. 2: Results for all 21 test cases of the first test presented in a histogram. The numbers from -105 to 105 indicate the angles in which the loudspeakers were. Black boxes indicate the loudspeakers that were emitting sound in each case. The height of the grey bars indicates how many times the particular loudspeaker was marked as emitting sound, 20 being the maximum.

The results of the K-S test indicated that most case distributions are significantly different from each other. However, particularly interesting are the cases in which the K-S test indicated a lot of similarity. Highest *p*-values are between cases 7&15, 6&15, 9&10, 10&17, 12&13, 14&17, and 16&18. All these case similarities, particularly between cases 6&15 and 7&15, strongly indicate that the resolution of spatial distribution perception is not adequately high for the subjects to be able to perceive gaps of 15° in the sound source.

In addition, the K-S test was used to analyze if any case distribution was close to a uniform distribution, i.e., a situation where subjects marked loudspeakers with equal probability. This hypothesis was rejected for all 21 cases, which proves that the subjects did not answer randomly.

Discussion

As mentioned, the test cases can be divided into four groups. The first group – cases from 1 to 7 in Fig. 2 – tested the auditory width perception. The cases with 1, 3 and 5 loudspeakers were perceived quite accurately. However, with more loudspeakers the furthermost loudspeakers were clearly less often marked, which indicates that a wide sound source was perceived a little narrower than it was. In addition, the middle loudspeakers were not marked as often as the loudspeakers in the edges on the far left and right. This indicates that it was challenging to tell apart whether all the loudspeakers between the perceived edges were emitting sound or not.

The second group – cases 8-13 – tested the perception of a gap in the sound source. Overall, the results indicate that it was difficult to accurately perceive a gap or its edges in the sound source. The tendency was to perceive an even wider gap than it actually was. When the situation is thought to be a setup with two wide sound sources it can be said that those sound sources were perceived narrower than they actually were. In addition, when these cases are compared to cases 1-7, it can be seen that the perception of a wide sound source is different when there are two of them presented simultaneously than when there is just one wide sound source.

The third group – cases 14-16 – consisted of chessboard-type combinations. These cases tested whether the subjects could perceive such complex sound source setups accurately or not. In cases 14 and 15, it is clear that the subjects' perception did not match the sound source composition. Rather, the subjects often marked the loudspeakers between the actual sound-emitting loudspeakers, which indicates that such narrow gaps in a wide sound source could not be perceived. In case 16 the two 30° wide sound sources on the right and left were perceived fairly accurately but the center area was inaccurate, as the center loudspeaker was often marked even though it was not emitting sound. As noted before, the statistical analysis indicates similarity between the perceptions of cases 7&15 and 6&15. Even though only every other loudspeaker was emitting sound in case 15 and every loudspeaker in cases 6 and 7, the subjects perceived them quite similarly. Thus, it is clear that a gap of 15° was not perceived.

Cases 17-21 tested the resolution of perception of three sound source groups - or alternatively, the perception of two holes. The results are in correspondence to the cases 8-13 with two sound source groups – the perceived width of the groups was mostly narrower than the actual width and conversely, the gaps were perceived wider than they actually were.

The average number of marked loudspeakers was, in most cases, smaller than the actual number of loudspeakers emitting sound. The highest average was 6.85 in case 7 where the actual number of loudspeakers was 13.

EXPERIMENT 2

The purpose of the second listening test was to further study the resolution of perception of a spatially distributed sound source. The first listening test indicated that the resolution of perception was not adequately high to perceive the smallest details of the sound source distributions in the test. Consequently, the desire was to further investigate the phenomenon and determine more precisely the accuracy of spatial distribution perception using different loudspeaker setups with various samples. Moreover, the effect of bandwidth was studied in order to find out if the spatial distribution is easier to perceive with wide-band noise than with narrow-band noise.

Experimental setup

The experimental setup in the test was the same as in the first test (illustrated in Fig. 1) with the exception that this time there were only 13 loudspeakers surrounding the listener. However, as the loudspeakers that were farthest on the left and right were not actually emitting sound in the first test, the setup can be considered to be similar in both tests.

Stimuli

A total of 13 different samples were used. They can be divided into three groups: bandpass filtered noise, brown noise and sinusoid samples. Again, all samples were 1000 ms in length - a 100 ms fade-in, 800 ms constant loudness and a 100 ms fade-out. Each bandpass filtered noise sample and brown noise sample had 13 uncorrelated channels.

The main focus was on the bandpass filtered noise cases. Two center frequencies, 500 Hz and 4000 Hz, and five filter bandwidths, 1, 1/3, 1/8, 1/12 and 1/24 octaves in width, were selected as the parameters, resulting in 10 samples. In addition, there were two different sinusoid cases where sinusoids with seven different frequencies were selected so that the frequency range was similar to the octave-band noise cases, i.e., from 353.5 to 707.1 Hz and from 2828.4 to 5656.9 Hz. Inside these ranges, the sinusoids were separated with equally large intervals.

Five loudspeaker setups were selected, with 2, 3, 4, 5 and 7 loudspeakers emitting sound. All setups included the leftmost and rightmost loudspeakers, and the others were selected so that they were evenly divided, i.e., each loudspeaker was equally far from one another in the pattern. Four different pairs were formed from the five loudspeaker setups: 2&3, 3&4, 4&5 and 5&7. The five loudspeaker arrangements are presented in Fig. 3.

Procedure and test hypotheses

In each test case, two different loudspeaker setups were presented on a touch screen. Then, after pressing ’play’, the subjects heard a sample as played back with those two setups. The task of the subjects was to discriminate which of the two loudspeaker setups was used in the latter sound event, i.e., the task was a two alternative forced choice procedure. The two sound events were played back twice with a short pause between them – as illustrated in Fig. 4 – and the subjects were forced to listen to the whole chain before answering. It was not possible to listen to the chain again. The subjects were instructed to face towards the center loudspeaker and remain still when the sounds were playing. There was a short training session prior to the actual test, after which the subjects heard each of the 52 cases four times, resulting in a total of 208 test cases.

Two main hypotheses are stated. First, the perception of spatially distributed sound sources becomes more challenging as the number of loudspeakers is increased. Second, the bandwidth affects in such a way that the perception becomes more challenging as the bandwidth becomes narrower.

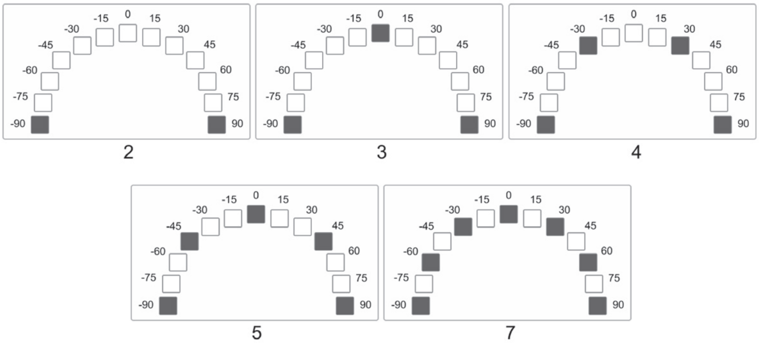

Fig. 3: The five loudspeaker setups of the second test. The black boxes represent the loudspeakers that were emitting sound. The numbers by the boxes indicate the angles in which the loudspeakers were.

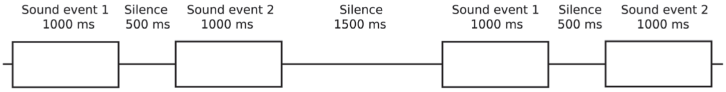

Fig. 4: Timeline presentation of one test case. The durations of sound events and silences are in milliseconds.

The cases with a center frequency of 500 Hz are in the frequency region where interaural time difference (ITD) dominates localization, whereas the cases with a center frequency of 4000 Hz are in the frequency region where interaural level difference (ILD) dominates localization. Thus, the effects of different localization cues may be studied.

A total of 19 subjects participated in the test. Again, they all were staff or students in the Department of Signal Processing and Acoustics of Helsinki University of Technology, and the authors did not participate in the test. Some of the subjects had participated in the first test as well. Almost half of the subjects had never before participated in any listening test, so they can be said to be naive listeners. None reported any hearing defects.

Results and analysis

The combined results and their 95% confidence intervals for all the subjects are presented in Fig. 5. There, the cases are divided into three subfigures: on the left are the cases with center frequency of 500 Hz and 4000 Hz, and on the right are brown noise and sinusoid cases. The different loudspeaker setup pairs are presented with individual lines in all subfigures. It should be noted that the points of each sample are horizontally misaligned only for easier visual inspection – bandwidth was always the same inside each group.

As the task was a two alternative forced choice procedure, the probability of answering correctly by pure chance is 50%. Commonly, the threshold for determining whether the answers were correct by knowledge or chance is 75%. In other words, any case where the percentage of correct answers is less than 75% cannot be said with confidence to be perceived accurately.

The results were statistically analyzed using a three-way analysis of variance (ANOVA). It was used to test which of the test variables – sample, loudspeaker setup pair and test subjects – had a statistically significant effect on the results. The results of the ANOVA indicated that loudspeaker setup pair clearly had the most statistically significant effect. The test subject had a statistically significant effect as well, but the sample did not.

As can be seen in Fig. 5, the cases with 2&3 loudspeakers were perceived most accurately with all 13 samples. In contrast, the perception accuracy of all the other loudspeaker cases was always below the threshold of 75%. However, the confidence intervals of the cases with 2&3 loudspeakers stay above 75% only with five of the samples. Nevertheless, it can be said that the results suggest that the first hypothesis is true – increasing the number of loudspeakers made the perception more difficult. This is supported by the above-mentioned ANOVA results and Tukey’s Honest significant difference test, which indicates that the loudspeaker setup pairs can be divided into three significantly different groups: 2&3 loudspeakers in the first, 3&4 in the second, and 4&5 and 5&7 in the third group.

The effect of bandwidth can only be examined in the region where ITD dominates localization, i.e., with the center frequency of 500 Hz. There, the results of the cases with 2&3 loudspeakers show that as bandwidth became narrower, the perception accuracy decreased, as only with the two widest bandwidths the confidence intervals were above the threshold of 75%. When these five cases were analyzed separately, the ANOVA indicated that both the sample (i.e., bandwidth) and the test subject had a significant effect on the results. This suggests that when the discrimination accuracy was above the threshold of 75%, the bandwidth of the noise signal did have a statistically significant effect. In the ILD region, i.e., with the center frequency of 4000 Hz, the effect of bandwidth cannot be examined for the discrimination accuracies were under the threshold of 75%. In conclusion, further studies are needed on the effect of bandwidth.

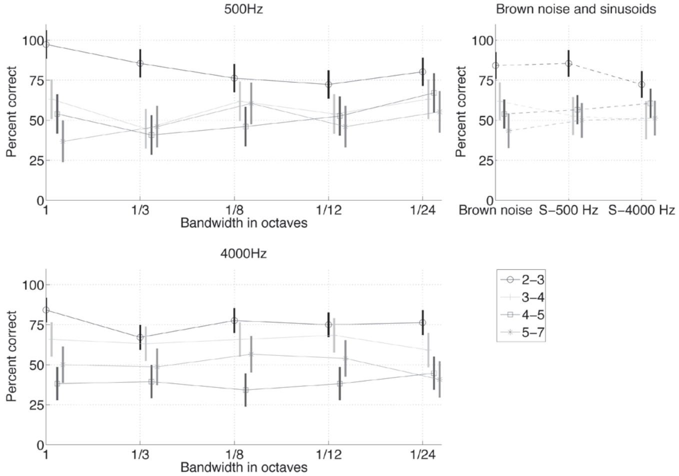

Fig. 5: Results of the second listening test. The percentage of correct responses in each case is presented in the y-axis. The different loudspeaker setups are presented with individual lines. S-500 Hz and S-4000 Hz stand for sinusoids around 500 Hz and 4000 Hz.

Discussion

Overall, it can be said that the discrimination task presented to the subjects was challenging. The perception accuracy for the cases with 3&4, 4&5 and 5&7 loudspeakers was below the threshold, which indicates that the resolution of spatial distribution perception was rather poor in this task. As for the effect of bandwidth, the results were less informative as was desired.

The sinusoid cases provide an interesting additional aspect to the results. As the frequency range in those cases was the same as with the octave band noise cases, their results can be compared. The results in the ITD region, i.e., in the cases with the center frequency of 500 Hz, are quite similar – the case with 2&3 loudspeakers was perceived accurately both with the octave band noise and the sinusoid case whereas the other loudspeaker pairs were not. In the ILD region the results look different. The case with 2&3 loudspeakers was perceived accurately with octave band noise but was not perceived correctly with the sinusoid case. This suggests that the frequency range is not the only attribute that affects the perception of these samples.

The bandpass signals can each be thought to be equivalent to a sinusoid that is modulated with a noise signal with a certain spectrum. It is assumed that the modulation causes differences in the signal so that when several loudspeakers emit uncorrelated signals, different loudspeakers become more audible, i.e., stand out at different times. With the octave band noise this effect is more clear, thus making the perception easier. When the bandwidth becomes narrower the effect is diminished and the perception becomes more difficult. The situation is similar with the sinusoid cases – there is no such modulation aid with pure sinusoids that are not close to each other in frequency and the perception is more difficult than with the octave band noise.

ACKNOWLEDGEMENTS

This work has been supported by The Academy of Finland (project no. 105780) and by the Emil Aaltonen Foundation.

REFERENCES

Cabrera, D., and Tilley, S. (**2003**). “Parameters for auditory display of height and size,” in *Proceedings of the 2003 International Conference on Auditory Display* (Boston, MA, USA).

Hirvonen, T. (**2007**). “Perceptual and modeling studies on spatial sound,” Report 83. Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland.

Hirvonen, T., and Pulkki, V. (**2006**). “Perception and analysis of selected auditory events with frequency-dependent directions,” J. Audio Eng. Soc. **9**, 803–814.

Hiyama, K., Komiyama, S., and Hamasaki, K. (**2002**). “The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field,” in *The 113th AES Convention* (Los Angeles, California, USA).

Mason, R., Brookes, T., and Rumsey, F. (**2005**). “Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli,” J. Acoust. Soc. Am. **117**, 1337–1350.

Massey, F. J. (**1951**). “The Kolmogorov-Smirnov Test for Goodness of Fit,” Journal of the American Statistical Association **46**, 68–78.

Perrott, D. (**1984**). “Discrimination of the spatial distribution of concurrently active sound sources: Some experiments with stereophonic arrays,” J. Acoust. Soc. Am. **76**, 1704–1712.

Perrott, D., and Buell, T. (**1981**). “Judgments of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise,” J. Acoust. Soc. Am. **72**, 1413–1417.