II

Publication II

Hirvonen, T. and Pulkki, V., "Center and Spatial Extent of Auditory Events as Caused by Multiple Sound Sources in Frequency-Dependent Directions", Acta Acustica united with Acustica, Vol 92, No. 2, Jan 2006.

Center and Spatial Extent of Auditory Events as Caused by Multiple Sound Sources in Frequency-Dependent Directions

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

Summary

In everyday life, humans often perceive complex auditory events. Many natural sound sources are not point-like, but spatially extensive. Also, sound reproduction techniques sometimes produce virtual sources whose directional cues propose multiple directions at a time. Perceptual mechanisms for decoding the perceived direction and the spatial distribution of such auditory events are not well known. This paper investigates how sound objects whose localization cues indicate different azimuth direction as a function of frequency, 1) are localized, and 2) how horizontally wide they are perceived. A horizontally wide (45°) sound source was created by presenting spectrally consecutive, non-overlapping, bandlimited noise samples simultaneously from the different loudspeakers of a loudspeaker grid in an anechoic environment. The narrowband samples together formed a broadband stimulus. The order of the narrowband noise samples in the loudspeakers, as well as the total frequency range of the samples, was varied from case to case. In each test case, the subjects were asked to indicate the perceived center of gravity of the sound image, as well as the direction of all the loudspeakers that they perceived to radiate sound. Generally, the perceived center could not be predicted merely by a simple model using Raatgever's frequency weighting function for binaural salience [1]. Alternative frequency weights were calculated analytically from the listening test results. The results also indicated that the perceived width of the sound sources produced by the nine-loudspeaker setup was, in all cases, less than half of the actual width of the source. This implies that some frequency bands from different loudspeakers fused together spatially.

PACS no. 43.66.Qp, 43.66.Pn

1. Introduction

The purpose of this paper is to investigate the perceptual aspects of complex auditory events where the localization cues imply different azimuth directions as a function of frequency. Such complex events are nowadays commonplace, as complicated multichannel systems and spatial audio algorithms are in wide use. The research presented here was initiated by a previous study, in which subjects reported the perceived directions of noise stimuli created using various simulated microphone and spatialization techniques and multi-channel loudspeaker setups [2]. In the study, the localization cues produced by each stimulus had also been investigated via computational modeling. It was discovered that in situations where the implied direction of the cues changed as a function of frequency, or the cues contradicted one another, the perceived direction could not be predicted with simple auditory modeling methods, such as averaging cues over frequencies. It seemed that the subjects emphasized some frequencies more than others, and/or utilized higher-level perceptual processing in determining the perceived direction. Furthermore, based on verbal comments, some cases were perceived more horizontally wide than others. The perception of width was not accounted for in the utilized auditory model.

In this paper, we state hypotheses about how humans perceive the direction and spatial distribution of a horizontally wide sound source. The hypotheses are tested by conducting a listening test where the different frequency bands of a broadband sound source are simultaneously presented from different azimuth directions in an anechoic environment. Throughout the paper, it is important to distinguish between the physical and perceptual domains. "Source" (e.g. loudspeaker) refers to the former, whereas the perceived sound is denoted as "event", "object", or "image" [3].

1.1. Background

This section reviews known results that provide insight into the topic of this paper. It has been established that humans use a variety of localization cues to determine the direction of incoming sounds. This paper is limited to

Received 6 June 2005, revised 25 November 2005,
accepted 28 November 2005.

hearing in the horizontal plane, where Interaural Time Differences (ITDs) and Interaural Level Differences (ILDs) have been widely accepted as the most prominent of all localization cues [3]. The neural coding of these interaural cues is traditionally modeled to take place at the peripheral levels of the auditory system [4]. This coincides with the physiological evidence that ITD and ILD are mainly encoded within specific frequency channels in the Medial and Lateral Superior Olive, respectively [5]. The neural response patterns are then processed by perceptual mechanisms with feedback from the higher stages and possibly with some inter-channel interactions. These higher levels of the hearing system and their development are not well understood. What is known is that infants, at least to some extent, learn to associate the neural activity patterns related to certain localization cue values to specific directions as they grow up [6]. As a result, a single point-like, broad-band sound source in an anechoic environment usually produces cues that are strongly associated with the actual left/right direction of the source across the utilized frequency range. Such cues are therefore noted to be consistent over frequency.

It has been found that when both ITD and ILD are consistent over a wide frequency range, but conflictingly indicate different source directions, the low-frequency ITD cues dominate localization [7]. The importance of low-frequency ITD is often attributed to the synchrony between the sound waveform and neural output, which enables the waveform of the signal to be coded accurately [8]. This is why the low-frequency ITD is sometimes called waveform ITD. At higher frequencies, on the other hand, the neural synchrony begins to decline and the ITD of only the signal envelope can be detected accurately. In the case where either ITD or ILD is set to be inconsistent as a function of frequency, thus giving conflicting directional information, the consistent cue is more prominent [9]. The case in which both cues are inconsistent and their implied directions vary notably over frequency has not been studied thoroughly. An open question is whether some frequency regions are perceptually more prominent than others in these situations.

In this respect, the research in this paper is closely related to the “dominant region” experiments by Raatgever and Bilsen [1] [10]. The purpose of their studies was to extract a perceptual weighting function for the salience of binaural components as a function of frequency. The tests were performed so that three frequency bands were presented to subjects using headphones. Initially, the middle-frequency band had a larger ITD value than the other two bands, and the subjects adjusted the amplitude of this band until the subjective lateral direction of the auditory event was the same as when the two other bands had the larger ITD value. Although the tests were performed only below 1200 Hz, a dominant region was located, centered around 600 Hz. Thus the results imply that some frequencies might contribute to the localization of complex sounds more prominently than others when the cues vary as a function of frequency. Raatgever’s results were used by Stern et al. to calculate a third-order polynomial approximation of the perceptual weighting function that was utilized in their auditory model [11]. Due to the lack of experimental data, Stern’s approximation was implemented with a constant value above 1200 Hz. Figure 1 illustrates the weighting function in the range of 100–1200 Hz.

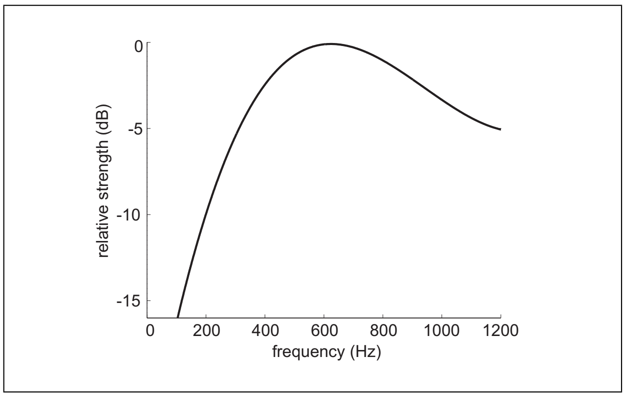

Figure 1. Approximation of perceptual weighting function for the salience of binaural components as a function of frequency based on measurements by Raatgever [1].

In addition to localization, this paper studies the perceived width of sound. The perceived width of broad-band sounds, whose interaural cues change unnaturally as a function of frequency has not, to our knowledge, been explicitly studied. The concepts of apparent source width (ASW) and perception of auditory spaciousness are often used to describe the perceived width. ASW is defined as the perceived horizontal spatial extent of the auditory event. In a normal listening room ASW and spaciousness depend on attributes such as the amount of early lateral reflections [12]. There are some studies that indicate how ASW can be manipulated also in anechoic conditions, such as used in these tests. Early research on width, or tonal volume, showed it to be a function of loudness, duration of the sound, interaural characteristics, and frequency [13] [14]. Potard and Burnett state that low-frequency sounds tend to have a larger apparent width [15]. The perception of spaciousness has been observed to increase with decreasing interaural cross-correlation (IACC) [16], a measure for similarity between the ear input signals.

The size of the physical sound source itself can vary and it understandably affects the perception. A seashore, for example, can be thought of as an extremely wide sound source. The question then arises as to whether the sound of a seashore actually consists of several smaller physical sources and is it perceived as one or as several sound objects. Thus, the concepts of sound fusion and segregation although usually ignored, are also relevant when studying ASW. Segregation here refers to perceiving separate sound objects with, for example, different frequencies or directions. According to Gardner, the same attributes that mediate ASW can also be responsible for sound segregation [17].

For the sake of simplicity, the perceived center of gravity of can be used to describe the localization of multiple-object images. This refers to weighting the directions of different objects of the event according to their loudnesses in order to determine a single center direction. Similarly, one can consider only the overall width, not taking segregation nor fusion into account. When several small physical sound sources, such as loudspeakers, are simultaneously active, the concept of ensemble width is used to describe the overall perceived width [18].

It should be noted that many of the listening tests discussed above were conducted using headphones. Headphones allow for separate control of both ear input signals and are thus suitable for theoretical experiments. However, auralization with headphones is prone to significant inaccuracies [19]. Headphones are known for their in-head localization, caused by unsuitable reproduction methods, the lack of head rotation cues, spectral distortions, and other factors [20]. It can be argued that it is preferable to use loudspeakers placed in an anechoic chamber when accurate reproduction and directional precision are needed, as is the case in our experiments. Of course, there are several localization studies that have employed loudspeakers placed in an anechoic environment. Their focus, however, has been on the localization of a single sound source. Perceived width in the absence of reflections has not, to our knowledge, been widely studied apart from the effects of IACC.

1.2. Hypotheses

The previous section discussed the relevant past research so that the hypotheses for the present investigation can be stated: When an auditory event whose auditory cues vary with frequency is presented to a listener,

1. The perceived center of gravity of the image can be predicted by considering the perceptual weights given by the Raatgever frequency weighting function.
2. The perceived horizontal extent of the event is equal to the actual physical width of the corresponding sound source.

Additionally, the results are examined from the viewpoint of perceptual segregation. We analyze whether the auditory events consist of one or several clearly separable spatial sound objects.

2. Methods

The experiments in this paper were conducted using a method where spectrally consecutive, but non-overlapping, narrowband noise samples are played through loudspeakers that were placed in different azimuth directions in anechoic conditions. Thus, the overall sound from all loudspeakers formed the stimulus in each test case. This can be interpreted as a horizontally wide sound source where the implied direction changes as a function of frequency; i.e., the produced interaural cues are generally inconsistent. However, within the narrow frequency ranges applied to each loudspeaker, ITD and ILD imply the same direction and are not in conflict with each other. In the experiments, the sound from the loudspeakers arrived to the listening position at the same time, eliminating the precedence effect. Also, the IACCs of the stimuli were high within each loudspeaker band, so its effect is also left outside this research. All loudspeakers were active in all test cases, and thus the physical width of the sound source was always constant during the tests. Only the utilized frequency range and the mutual order of the frequency bands in the loudspeaker setup were varied.

2.1. Experimental Setup

A nine-loudspeaker setup shown in Figure 2 was constructed in an anechoic chamber with a lower frequency limit of 100 Hz. The loudspeakers (Genelec 1029A) were mounted to a linear bar approximately 2 meters away from the listener and their distances were compensated for by the use of delays to be equal within one centimeter. This together with careful listener positioning ensured that the signal from any one speaker would not arrive before others and cause precedence effect, which would bias the perceived direction strongly [21]. Further research on temporal effects on complex auditory events is left to future.

The loudspeakers covered the azimuth sector symmetrically from −22.5° to +22.5° at the approximate height of the subject's head when in the listening position. Thus, the angular interval between adjacent loudspeaker centers was 5.6°. The loudspeakers were visibly labeled with numbers 1–9 from left to right.

In addition to the loudspeaker distance alignment with delays, equalization of the test system magnitude response was required. Even though the free-field response of each loudspeaker is close to flat above 60 Hz, placing nine speakers in close proximity introduced unwanted effects on their frequency responses due to diffraction. To correct the situation, the magnitude responses of all loudspeakers in the listening setup were measured and a real-time, FFT-filtering-based equalization was implemented. The resulting loudspeaker magnitude responses were measured and found to be flat within 1.5 dB in the region of 0.1–7 kHz.

2.2. Procedure

To test the stated hypotheses, a listening test composed of two tasks was arranged. The subjects listened one at a time to the test cases in the anechoic chamber. For each stimulus, they were asked to indicate 1) the loudspeaker closest to the center of the sound, and 2) all loudspeakers that seemed to radiate sound. These two tasks were performed in separate sessions. The order of the test cases and the two tasks were randomized for each subject. The subjects performed the listening tests seated on a chair facing the middle loudspeaker of the listening setup. The subjects were instructed to use a provided headrest to remain in the correct position during the evaluations. The subjects' responses were registered using a keyboard, whose number keys from 1 to 9 indicated the loudspeakers of the listening setup.

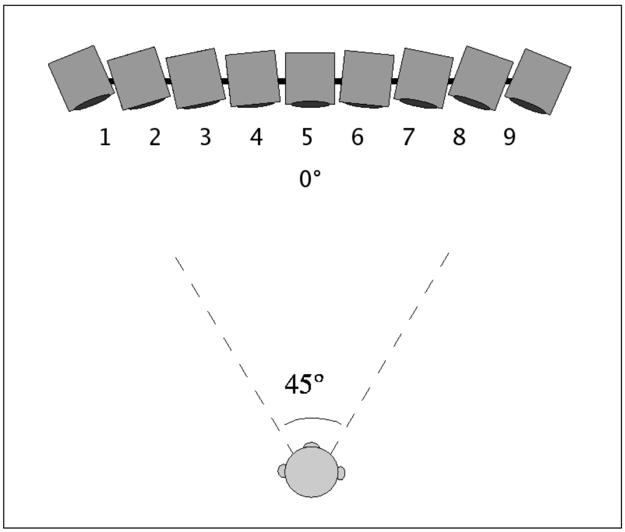

Figure 2. The listening setup used in the tests. Nine numbered loudspeakers were suspended vertically at the height of the subject's head. Loudspeaker distances to the listening location were compensated with delays so that the sound from all loudspeakers arrived at the listening position at the same time. The loudspeakers covered the azimuth sector symmetrically from -22.5° to 22.5°.

Each test case consisted of a noise sequence of 1 s length, followed by a silence of 400 ms. The noise sequence had a 20 ms fade-in and fade out, and a random Gaussian envelope. The resulting 1400 ms segment was looped for the duration of the time it took for the subjects to give their responses, after which the next test case was presented. The subjects could also correct any erroneous key strokes. By using repeated one-second samples, we wanted to examine the perception of signals that were more of a steady-state nature than impulsive. The 400 ms pauses were created to ensure that the short-term echoic memory would reset [22].

In one test session, the subjects were instructed to choose one loudspeaker that they judged to be in the middle of the sound they heard by pressing the corresponding number key. In cases in which the subjects would perceive horizontally wide or multiple-image sound events, they were instructed to estimate the center of gravity of the sound. In the other session, the subjects could press any combination of keys corresponding to the loudspeakers perceived to produce sound. The subjects had been informed that the number of perceived speakers could be anything between one and nine. Due to the discrete nature of the responses in both tasks, the azimuth resolution of the experiments was approximately 5.6°, i.e. the distance between the loudspeaker centers. The subjects were familiarized with the stimuli and the test system before each session by playing a few random samples, as well as by explaining the task they were asked to do.

The method described above bears some similarities to the source-identification method discussed by Hartmann [23]. It involves identifying the location of the sound when one of several visible sound sources is radiating sound. Hartmann used a loudspeaker grid from which the subjects chose one as the probable source [24]. This procedure yielded ample accuracy and thus it was deemed suitable also for this research.

2.3. Stimuli

The Equivalent Rectangular Bandwidth (ERB) scale [25] was used as a basis for dividing the stimulus into narrower frequency bands that were routed to the different loudspeakers of the test setup. The ERB scale was deemed suitable for these experiments, since one ERB describes the auditory resolution in the frequency domain when listening to broadband sounds.

All loudspeakers were active in all test cases and the bandwidth of one speaker was either 1, 2, or 3 ERBs, depending on the case. In order to obtain the bandlimited loudspeaker signals, a Gaussian noise signal in the range of 100–5858 Hz was split into 27 consecutive ERB bands in the frequency domain. The filtering was performed using FFT; the signal was Fourier-transformed, multiplied with a rectangular filter window corresponding to the desired frequency band, and inverse-transformed back to the time domain. Thus, the narrowband signals have zero correlation with each other. The relative loudnesses of the narrowband samples were carefully adjusted with subjective listening by several persons to be as identical as possible. As the loudspeaker magnitude responses and their relative levels had also been aligned, all speakers had equal perceptual weights in terms of loudness and bandwidth in every test case.

We opted to limit the present investigation to nine basic test cases, that were repeated in different frequency regions. In the fundamental case, the frequency increases in a stepwise manner with increasing azimuth angle. The eight remaining cases were created by cyclically rotating the nine frequency bands in the loudspeaker setup. In this context, cyclical rotation means that the noise band sample in the rightmost loudspeaker was moved to the leftmost speaker and other bands were shifted one speaker to the right. The spatial frequency configurations of the test cases are illustrated in Figure 3 in the following section. The purpose of these test cases was to reveal how different frequencies dominated the perceived direction as the same narrowband signals were presented from different directions.

The concept of test scheme is used in this paper to indicate the utilized frequency range. We have used nine schemes, all of which contain the above-mentioned nine cases. The schemes are named according to which of the 27 ERB-band signals were used, 1 indicating the lowest and 27 the highest frequency band. Initially, we chose five test schemes; these were to use ERB-bands 1–9, 10–18, and 19–27 for cases where the bandwidth of each loudspeaker was one ERB, as well as ERBs 1–18 and 1–27 for two- and three-ERB loudspeaker bandwidth cases, respectively. In addition to the previous test schemes, four additional schemes were examined using a smaller number of subjects. We wanted to investigate the ITD range more carefully and thus utilized three test schemes with ERB-bands 4–12, 6–14, and 8–16. The high-frequency scheme 19–27 was also tested "inverse". This means that instead of moving cyclically from left to right, the bands were oriented as a mirror image so that the frequency increased when moving from right to left in the loudspeaker setup. The bandwidth of each loudspeaker in the additional test schemes was also one ERB. Table I summarizes the test schemes used in this study.

A noteworthy characteristic of the test cases is that most have a distinctive "discontinuity point", where an abrupt change in frequency occurs within a small spatial angle. In the first case in each scheme, the frequency increases in small steps when moving from left to right (or right to left) in the loudspeaker setup. In the other eight cases the lowest and the highest frequency bands are presented from adjacent loudspeakers.

To summarize, the test included nine test schemes with different frequency regions, each of which consisted of nine test cases. Thus, there were altogether 81 test cases.

2.4. Test Subjects

A total of fourteen subjects participated in the tests. However, some test schemes had only five participants due to lack of resources. All subjects were students or staff in the Laboratory of Acoustics and Audio Signal Processing of Helsinki University of Technology and aged between 20 and 40 years. Although at least some of the subjects had notable musical skills and experience in listening tests, none had experience with this specific task. None reported any hearing defects.

3. Results

3.1. Perceived Center

In the remainder of this paper, "perceived center" is used to refer to the mean subjective judgments of the task where the subjects indicated the perceived center of gravity of the sound events. Figure 3 presents the listening test results for the perceived center in all test cases. Each of the nine test schemes is depicted in its own panel of the figure. The title of each panel indicates the ERB-bands used in the scheme, as well as the utilized frequency range. Each of the nine rows on the y-axis represents one test case. The nine loudspeakers in the listening setup are represented on the x-axis. The different test cases in each scheme are represented with the darkest color box symbolizing the lowest, and the lightest one the highest frequency in each scheme. For example, the center frequencies of the lowest ERB-band samples in schemes 1-9 and 6-14 are 119 Hz and 368 Hz, respectively, although both are indicated by the same color. The mean of the subjects' responses is illustrated with a line in each panel and error bars give 95% confidence intervals for the means. No scaling or normalization has been applied to the results.

Table I. Test schemes and their respective frequency regions and stimulus bandwidths in each of the nine loudspeakers. (For example, if the bandwidth in each loudspeaker was 3 ERB, total stimulus bandwidth was 27 ERB). Each test scheme consisted of nine test cases in which the order of the loudspeaker frequencies was permutated differently.

test scheme	frequency region (Hz)	stimulus bandwidth per loudspeaker (ERB)
1–9	100-640	1
4–12	226-974	1
6–14	336-1264	1
8–16	472-1625	1
10–18	640-2072	1
19–27	2072-5858	1
19–27inv	2072-5858	1
1–18	100-2072	2
1–27	100-5858	3

Let us first examine the results for schemes 1–9, 4–12, 6–14, 8–16, and 10–18 in Figure 3. Here, the total utilized frequency range of the stimuli increases from 100–640 Hz to 640–2072 Hz. The first four schemes have frequencies mainly in the range of the waveform ITD cue, while the 10-18 scheme also contains frequencies where ILD is prominent as well. The perceived center often follows the direction of the discontinuity region where two adjacent loudspeakers produce ERB-band samples wide apart in frequency. In many cases, the perceived center is located within the sector of these two loudspeakers (span of circa 11.2°) or near it. This phenomenon is discussed in more detail in Section 4.

In contrast, when examining the results for the higher frequency region 2072–5858 Hz (scheme 19–27), it can be seen that the subjects have chosen loudspeakers 6 or 7 as the perceived center in almost all cases. This general bias to the right of the center of the loudspeaker setup might be caused by unequal perceptual weighting of the different ERB-bands, and/or by some unknown phenomenon, such as handedness or ear dominance. The latter has been shown to be a significant phenomenon for example when IACC is low [26]. To investigate this result further, a test scheme in which the order of the bands was inverted was also designed. This scheme is titled 19-27inv in Figure 3. When comparing these results to the non-inverse case, all case means except one exhibit a shift to the left. However, the mean curves in the two test schemes are not mirror images, so we cannot totally exclude the possibility of bias towards right caused by, for example, handedness. The reasons for the bias at high-frequencies are not clear at this point and more research is needed. The phenomenon is therefore not discussed further in this paper.

Finally, the results in which the bandwidths of each loudspeaker were two and three ERB-bands (test schemes 1–18 and 1–27) are presented in the middle and rightmost panels of the bottom row of Figure 3. In the two-ERB cases, the perceived center of the auditory event is focused on the middle of the setup. The three-ERB scheme (1–27)

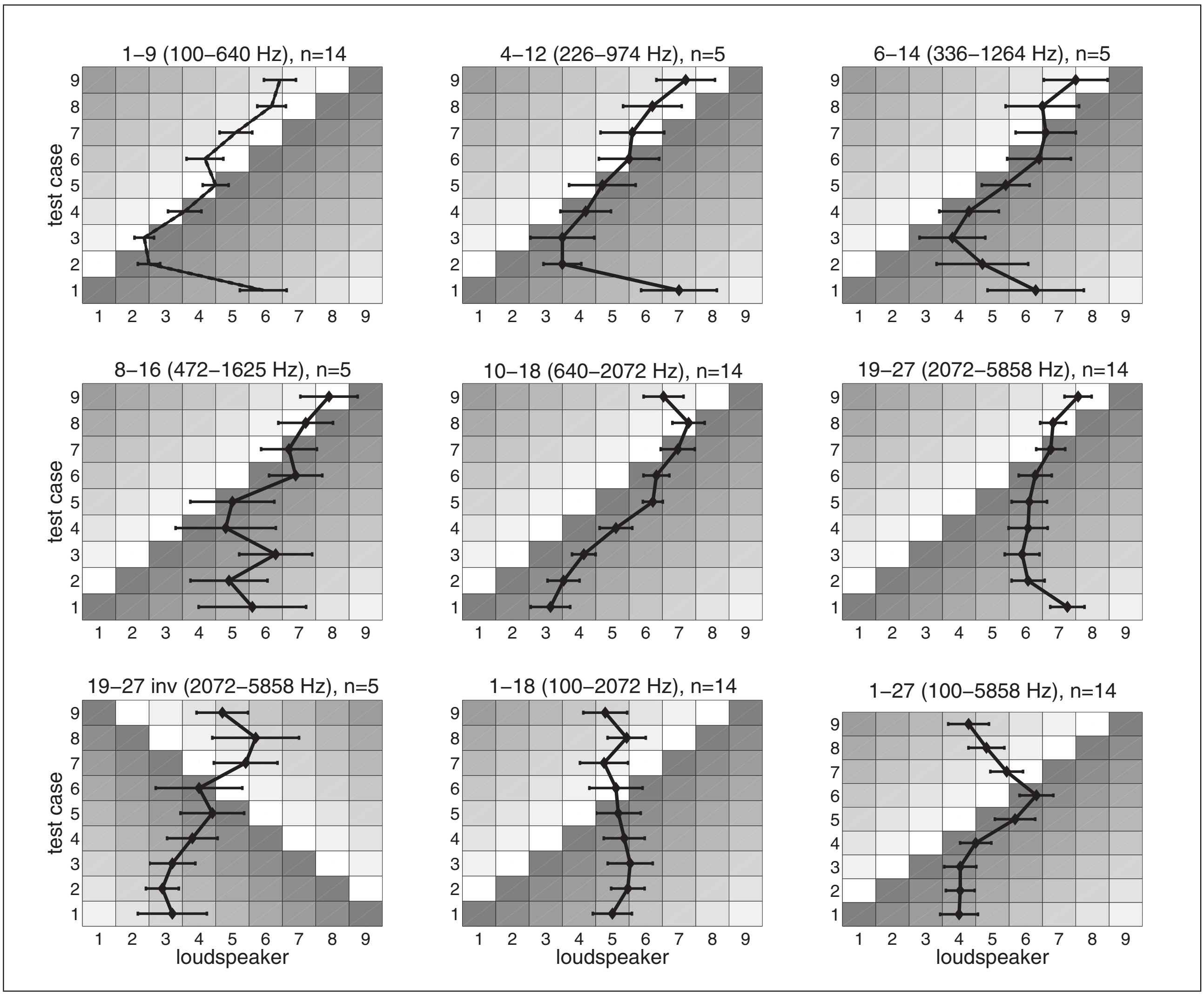

Figure 3. Results for perceived center of sound. Listeners indicated the center of gravity of a horizontally wide sound source that consisted of an array of nine loudspeakers. The stimuli consisted of narrowband noise samples played back from different loudspeakers. All loudspeakers were active in all test cases. The title of each panel indicates the ERB-bands used in that test scheme and the corresponding frequency range, as well as the number of test subjects (n). Colors from dark to white denote the utilized ERB-bands from the lowest frequency to the highest within the scheme's frequency range. The means of perceived centers are represented by dark lines with error bars giving the 95% confidence intervals.

is somewhat different; the mean response closely follows the discontinuity point in cases 4–6. Generally however, the results for the cases with two and three ERB-bands per loudspeaker seem to display less correlation with the direction of the discontinuity frequency region.

3.2. Perceived Width

In addition to indicating the perceived center of each stimulus, the listeners singled out the loudspeakers that in their opinion radiated sound. The individual distributions of the marked loudspeakers in each of the 81 individual cases were examined by the authors. The distributions are not plotted in this paper since different schemes had different numbers of subjects, which in turn would cause the plots to be misleading. In examination it was found that the indicated speakers generally surround the mean perceived center symmetrically, i.e. the perceived center was located in the middle of the extended auditory event. The deviations of indicated loudspeaker locations between individuals were not significantly larger than between the repetitions of the same subjects in each case, which indicates that the subjects perceived the sound events in a similar manner.

Figure 4 shows the mean perceived number of loudspeakers in each case, disregarding the actual direction of the indicated speakers. The nine cases of each test scheme are presented in their own panels. No scaling or normalization has been applied to the results. In addition to the perceived number of loudspeakers, a further measure is employed in Figure 4, i.e., "Width". This measure indicates the azimuth span between the leftmost and rightmost borders of a auditory event. Here, width is expressed in degrees and it is assumed that the span of one loudspeaker that was perceived as radiating sound was 5.6°, the approx-

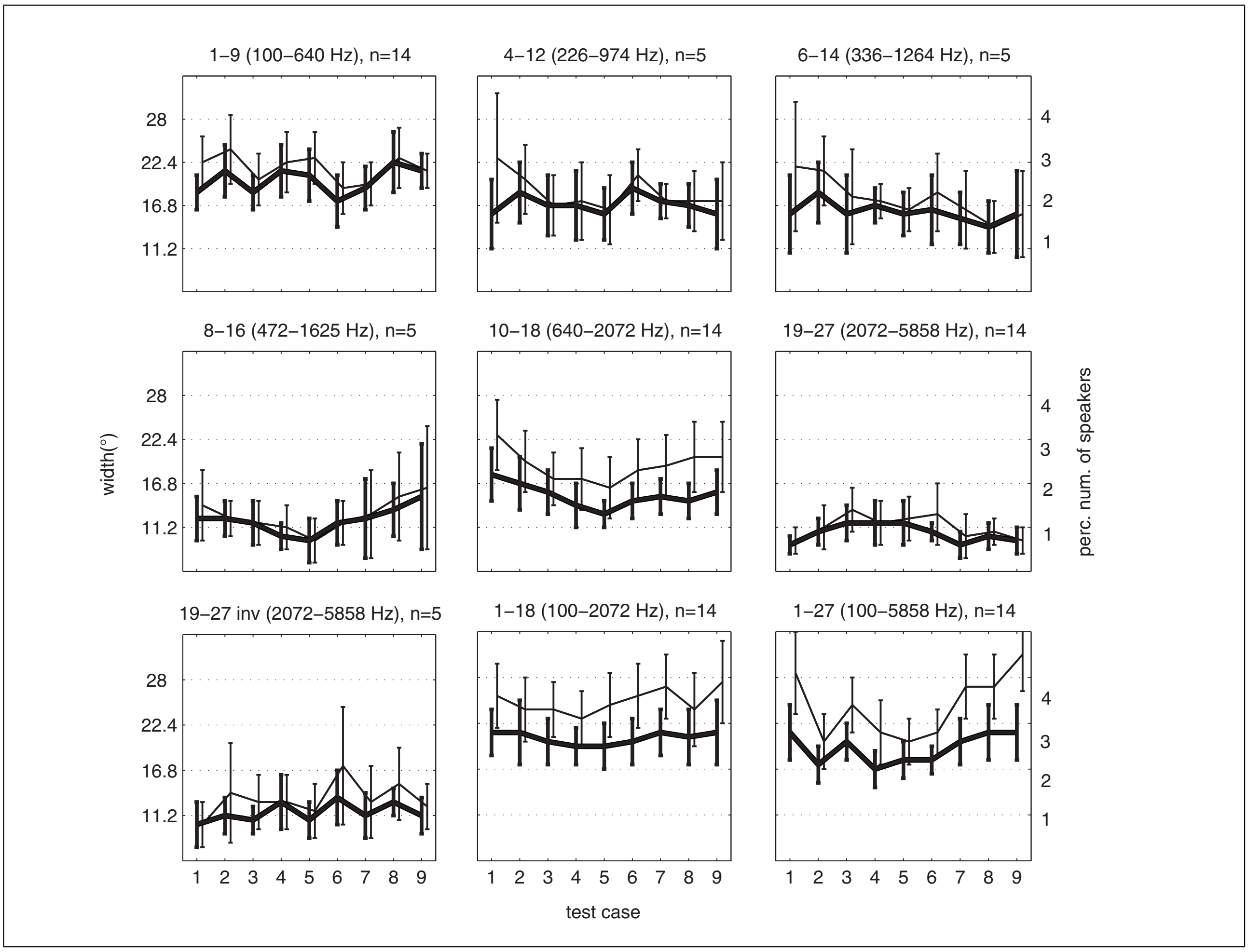

Figure 4. Mean perceived number of loudspeakers (thick line) and perceived width (thin line) in each test case. The title of each panel indicates the ERB-bands used in that test scheme and the corresponding frequency range, as well as the number of test subjects (n). The nine cases of each scheme are presented in the x-axis of each panel. Case numbers indicate the same test cases as shown in Figure 3. Error bars indicate the 95% confidence intervals of the data.

imate sector of one speaker in the test setup. The difference between the width and perceived number of loudspeakers is that the former does not take into account the gaps that were possibly present in the results. For example, if a subject marked the loudspeakers 3, 4 and 6 as radiating sound in one particular case, the width value would be 4×5.6° = 22.4° and the perceived number of loudspeakers 3. In this sense, the concept of width used here is similar to the "ensemble width"-attribute, rather than ASW or similar measures.

In all nine test schemes, the results show no significant differences between the different cases of the scheme beyond the confidence intervals. Thus, the present data proposes that the number of indicated speakers was not significantly affected by changing the azimuth location of the ERB-band noise samples in the loudspeaker grid.

3.2.1. Perceived Width and Spatial Segregation in Different Frequency Regions

It is now investigated how the the different frequency regions of each scheme affect the perceived width. Figure 5 illustrates the mean perceived number of loudspeakers and width averaged over all test cases in a particular scheme. The 95% confidence intervals in Figure 5 are relatively small as a result of using the data from all cases in each scheme, and they allow for some significant phenomena to be noted. When examining the test schemes 1–9, 4–12, 6–14, and 8–16 that have frequencies mainly in the waveform ITD range, the perceived width decreases significantly when moving to higher frequencies. This coincides with the hypothesis by Potard and Burnett that perceived width tends to be inversely proportional to frequency [15].

It is also interesting to investigate if the listeners perceived one or multiple spatially separable sound objects. In Figure 5, it can be seen that three test schemes (10-18, 1-18, and 1-27) exhibit a significant difference between the average perceived number of loudspeakers and the mean width values. In practice, this means that the subjects left some speakers between the marked speakers unmarked, i.e. there were gaps inside the extended auditory event. The difference between these values suggests that two or more spatially separate sources were perceived instead of one.

What is common to the previous three test schemes is that they include frequencies strongly in the range of both

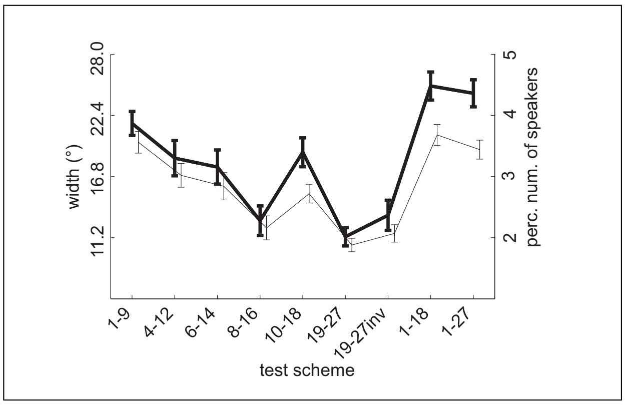

Figure 5. Mean perceived number of loudspeakers (thick line) and perceived width (thin line) for each test scheme. Error bars indicate 95% confidence intervals of the data.

waveform ITD and high-frequency ILD cues. It has previously been shown that humans can sometimes perceive separate images for these two cues in a complex stimulus [27]. This appears to also be the case here. When only frequencies above 2 kHz were utilized (test schemes 19-27 and 19-27inv), there were no significant differences between the width and the number of loudspeakers, and the perceived width was also rather low. Evidently the subjects perceived a single ILD-based spatial source in these cases.

The two- and three ERB-bands per loudspeaker cases show an increase in perceived width when compared to the low frequency scheme 1–9, although the number of marked loudspeakers is not significantly increased. It seems that extending the frequency range 100-640 Hz by adding high frequencies 640–2072 and 640–5858 Hz to the stimuli did not prominently increase the perceived width. The increase in the width value can be attributed to the fact that two or more separate sources were more likely perceived in these cases compared to the cases of the 1–9 scheme. Examination of the individual results showed that 32% and 39% of all width perception results with schemes 1–18 and 1–27, respectively, included gaps of one or more loudspeakers.

3.2.2. Coincidence between Physical and Perceived Width

It can be seen from Figure 4 that the 95% confidence intervals of means of perceived width never include values over 30°, and typically the means are between 10–20°. None of the test cases was ever perceived as radiating sound from all nine speakers. It could be argued that this phenomenon would result from the listeners' assumptions of the source width always being smaller than the width of the setup. If this was the case, it would mean that the test arrangement had introduced unwanted bias to the results. However, as a preliminary experiment, the authors had performed the same listening test and found the results to be similar to those presented here, despite the knowledge that all the loudspeakers were radiating sound. Thus it can be stated that the hypothesis 2 can be rejected in all tested cases since the widths of sound events were found to be about half or less of the actual size.

4. Modeling perceived center

To test hypothesis 1, i.e. if the perceived center could be estimated based on Raatgever's frequency weighting function, we examined it and two alternative methods with the listening test data. The analysis presented in this section is based on the assumption that the subjects did not utilize any form of interchannel perceptual processing between different ERB-bands. Rather, each ERB band is considered to have its own perceptual weight in the complex stimuli used in this study. This is done for simplicity, more complicated modeling techniques may be attempted in the future.

As mentioned in Section 1.1, Stern *et al.* have implemented a polynomial approximation for a binaural salience weighting function that is based on the measurements made by Raatgever [11]. This function is utilized here to calculate a prediction for the mean perceived center. Although Stern et al. used constant weighting above 1200 Hz, we have restricted our examination to test schemes 1–9, 4–12, and 6–14, since the frequencies of these schemes are limited to the range of Raatgever's measurements.

Predictions for the perceived center in each case are obtained by multiplying the azimuth direction of each frequency band with the relative weight value of the Raatgever function at the respective ERB-band's center frequency, after which a single predicted direction was obtained by summing the weighted azimuth angles:

$$\sum_{i=1}^{9} w_i * d_i = D, \tag{1}$$

where w_i is the relative normalized perceptual weight of the frequency band in direction d_i, and D indicates the predicted direction of the model. Weight values are normalized so that $\sum_{i=1}^{9} w_i = 1$ in each case.

The three upper panels of Figure 6 show the predictions of this simple method compared to the listening test results presented in Figure 3. The lower panels show the predictions with a modified weighting which will be discussed later. In each scheme, there are 3–7 cases where the direction estimated with the Raatgever function does not fall within the 95% confidence interval of the mean of perceived center. It can thus be concluded that hypothesis 1 must be rejected with this type of sound sources.

Since the weight values produced by Raatgever's function were not entirely appropriate here, it was in our interest to find more suitable alternatives.

4.1. Analytic Solving of Optimal Frequency Weights

We wanted to know the optimal weight values for each frequency band that would yield predictions similar to the subjective results when using the previously described simple model. The test arrangement allowed for solving

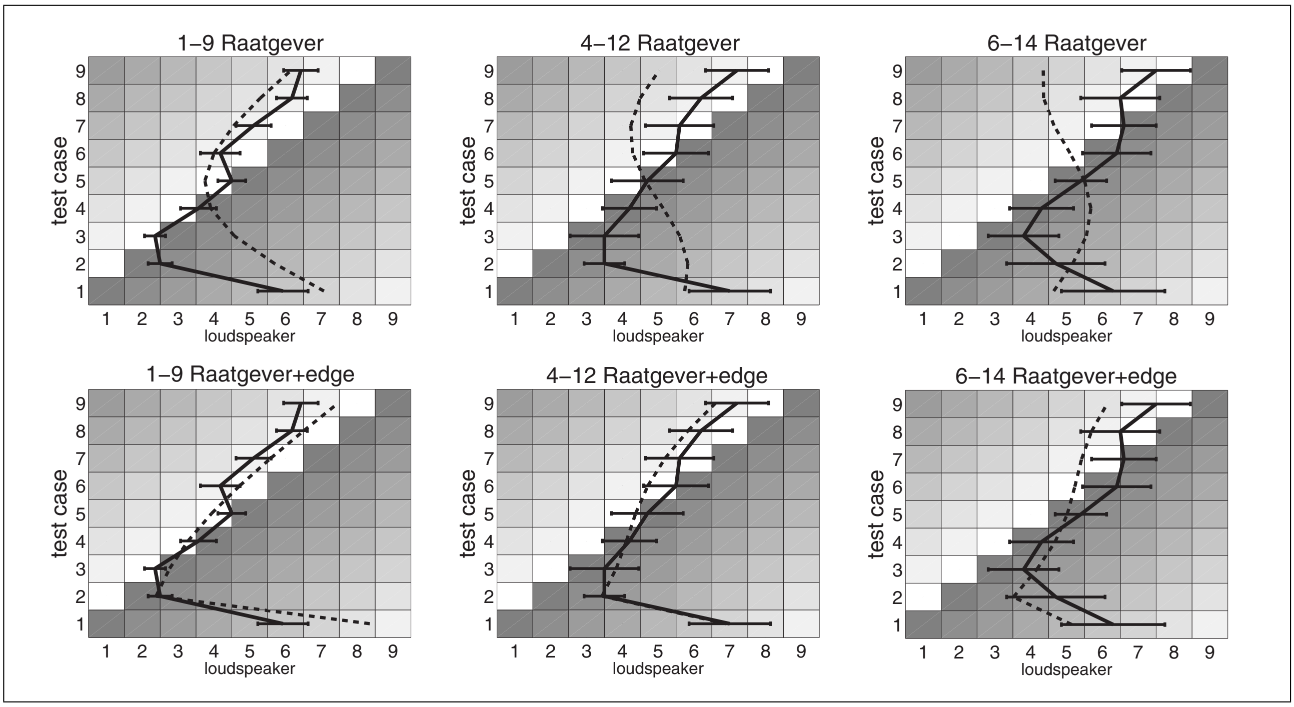

Figure 6. Upper row: comparison between the listening test results and the predicted center directions based on the binaural salience weighting function implied by the data of Raatgever [1]. Lower row: the model predictions fit the listening test data better when the "edge" bands of each scheme are given more weight. The weight values of the Raatgever function were multiplied by at the 5 lowest and by 10 at the highest band. Predictions for the perceived center in each case were obtained by multiplying the azimuth direction of each frequency band with the relative weight value at the respective ERB-band's center frequency, and summing the weighted azimuth angles. The solid line indicates the mean listening test results with 95% confidence intervals. The dotted lines show the model predictions.

these frequency band weight values analytically for each test scheme. As each test scheme consisted of nine test cases that utilized the same nine frequency bands, a system of nine linear equations was implemented. Each equation represented one case of a scheme: the spatial directions of the frequency bands were multiplied with unknown weight values and summed to obtain the subjective perceived centers. The nine optimal weight values for each of the nine frequency bands were then solved.

Figure 7 shows the results of these calculations. The nine weight values of each scheme are plotted in separate panels as continuous curves with a common frequency axis. Each value is located at the center frequency of the corresponding frequency band. Interestingly, a few of the weight values are negative. However, they are very close to zero. The Raatgever weight values that were used for the predictions shown in Figure 6 are also given for test schemes 1-9, 4-12, and 6-14. To ease comparison these values are normalized so that the sum of all weights in a scheme is 1.

In Section 3.1, it was established that the the lowest and the highest frequency band of each case seemed to play an important role when determining perceived center. This is also evident from the weight values in Figure 7. With all schemes in which the loudspeaker bandwidth was one ERB, the greatest weight is given to either the highest or the lowest band of the scheme. Interestingly, the maximum weight seems to shift gradually from the highest to the lowest band when moving from low (scheme 1-9) to higher frequencies (scheme 10-18). With schemes entirely in the ILD region (19-27, 19-27inv), or where the loudspeaker bandwidth is 2 or 3 ERB, the weight values are more even and it is harder to establish any clearly dominant frequency band: although both utilize the same frequencies, the greatest weight is given to the highest and the lowest band with schemes 19-27 and 19-27inv, respectively.

4.2. Modified Raatgever Frequency Weighting

The results in previous sections suggest that the lowest and the highest bands of the frequency content gain saliency in most test cases compared to the Raatgever function. It is now investigated if the model for perceived center can be extended based on this finding. A model was composed where the Raatgever function is used as the basis and the lowest and the highest bands of each scheme are given more weight. The lower three panels of Figure 6 present the center direction predictions when the weight values of the Raatgever function were multiplied by 5 at the lowest and by 10 at the highest band.

The dominance of both the lowest and the highest frequency band in the localization task requires further research to be fully understood. A possible explanation could be that in many cases the two edge bands together created a perceptually significant discontinuity, where the frequency "jumps" from low to high within a small spatial

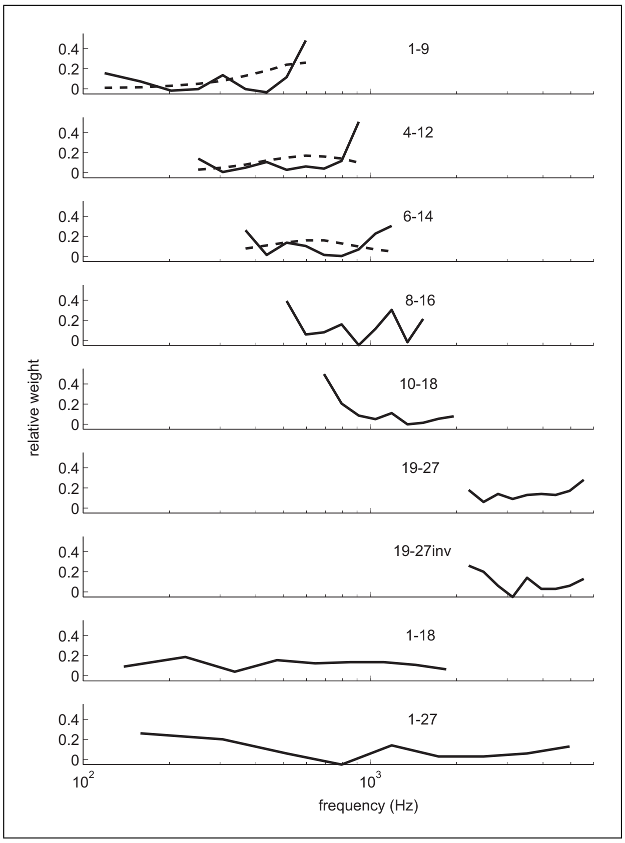

Figure 7. Optimal weight values (solid lines) for the frequency bands of each test scheme. The values are calculated from the listening test results for perceived centers with linear equations. Dotted lines show the normalized weights based on the Raatgever weighting function [1] for the three lower-frequency schemes 1–9, 4–12, and 6–14. The weights of each scheme are normalized so that their sum is 1.

sector. A strong contrast or discontinuity in, for example, color of a visual object contributes strongly to the perception in the visual domain. This kind of context-dependency implies higher-level processing that is not usually associated with localization, although some recent papers have proposed that this is also possible, for example [28].

5. Summary

The goals of this research were to investigate the perception of direction and perceived width of complex auditory events. A set of listening tests, in which ITD and ILD of horizontally wide sound sources were varied as a function of frequency, was conducted in an anechoic environment. For each sound, the subjects were asked to indicate the perceived center of gravity and all loudspeakers that, in their opinion, radiated sound. Generally, the results revealed several intriguing phenomena, most of which require more experiments to be adequately understood.

The main hypotheses for the investigation were stated as follows: First, the perceived center of the event can be predicted by considering the perceptual weights given by the Raatgever frequency weighting function. The second hypothesis assumed that the perceived horizontal extent of the event is equal to the actual physical width of the corresponding sound source.

Contrary to the first hypothesis, the perception of center direction could not be predicted by a model using the Raatgever weighting function for the salience of binaural components [1] that emphasizes frequencies around 600 Hz. Alternative weight values for different frequency bands were calculated analytically for each investigated frequency range based on the listening test results. These calculations indicate that it is difficult to establish any general dominant frequencies, but that the lowest and highest bands were perceptually important in all examined frequency regions. The exact reasons for this phenomenon are not clear and require further research.

The listening test results indicated that the perceived width of the auditory events produced by the nine-loudspeaker setup was in all cases less than half of the actual width. This contradicts the second hypothesis. This suggests that some frequency bands from different loudspeakers fused together spatially. No significant differences in width were found between the cases in which the total utilized frequency range was constant and the azimuth directions of different frequency bands were varied. It was therefore concluded that the main effects in perceived width were caused by the utilized frequency range. In the one-ERB bandwidth per loudspeaker cases limited to the ITD range, the perceived width increased as the utilized frequency range was lowered. Increasing the frequency range from nine to 18 or 27 ERBs by adding high frequencies above 640 Hz did not widen the auditory event prominently. When the range of the auditory event included frequencies both in ITD and ILD range, two or more horizontally separate sources were more likely to be perceived. This phenomenon also caused the overall width of the event to increase compared to the cases that were prominently in the frequency range of either cue alone.

Acknowledgments

The authors wish to thank Mr. Juha Merimaa for his many valuable suggestions. This work has been supported by The Academy of Finland (project # 105780).

References

[1] J. Raatgever: On the binaural processing of stimuli with different interaural phase relations. Dissertation. Technische H geschool Delft, 1980.

[2] V. Pulkki, T. Hirvonen: Localization of virtual sources in multi-channel audio reproduction. IEEE Transactions on Speech and Audio Processing **13** (2005) 105–119.

[3] J. Blauert: Spatial hearing. revised ed. The MIT Press, Cambridge, MA, USA, 1996.

[4] L. A. Jeffress: A place theory of sound localization. J. Comp. Physiol. Psych. **41** (1948) 35–39.

[5] J. O. Pickles: An introduction to the physiology of hearing. Academic Press, 1988.

[6] R. Y. Litovsky, D. H. Ashmead: Development of binaural and spatial hearing in infants and children. – In: Binaural and Spatial Hearing in Real and Virtual Environments.

R. H. Gilkey, T. R. Anderson (eds.). Lawrence Erlbaum assoc., Mahwah, New Jersey, 1997, 571–589.

[7] F. L. Wightman, D. J. Kistler: The dominant role of low-frequency interaural time differences in sound localization. J. Acoust. Soc. Am. **91** (1992) 1648–1661.

[8] A. R. Palmer, I. Russel: Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of the inner hair cells. Hear. Res. **24** (1986) 1–15.

[9] F. L. Wightman, 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, T. R. Anderson (eds.). Lawrence Erlbaum Assoc., 1997.

[10] J. Raatgever, F. A. Bilsen: A central spectrum theory of binaural processing. J. Acoust. Soc. Am. **80** (1986) 429–441.

[11] R. M. Stern, A. S. Zeiberg, C. Trahiotis: Lateralization of complex stimuli: A weighted image model. J. Acoust. Soc. Am **84** (1988) 156–165.

[12] M. Barron, A. H. Marshall: Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. J. Sound Vib. **77** (1981) 211–232.

[13] E. G. Boring: Auditory theory with special reference to intensity, volume, and localization. Am. J. Psych. **37** (1926) 157–188.

[14] D. Perrot, T. Buell: Judgments of sound volume: effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise. J. Acoust. Soc. Am. **72** (1982) 1413–1417.

[15] G. Potard, I. Burnett: A study on sound source apparent shape and wideness. Proceedings of Int. Conf. Auditory Display, 2003, 25–28.

[16] P. Damaske: Subjektive Untersuchungen von Schallfeldern (subjective investigations of sound fields). Acustica **19** (1967/68) 198–213.

[17] M. B. Gardner: Image fusion, broadening, and displacement in sound localization. J. Acoust. Soc. Am. **46** (1969) 339–349.

[18] F. Rumsey: Spatial quality evaluation for reproduced sound: terminology, meaning and a scene-based paradigm. J. Audio Eng. Soc. **50** (2002) 651–666.

[19] D. Hammershøi: Binaural technique: a method of true 3d sound reproduction. Dissertation. Aalborg University, 1995.

[20] C. A. Poldy: Headphones. – In: Loudspeaker and Headphone Handbook, 2nd ed. J. Borwick (ed.). Focal Press, 1994, 585–692.

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

[22] N. Cowan: On short and long auditory stores. Psychological Bulletin (1984) 341–370.

[23] W. M. Hartmann, B. Rakerd, J. B. Gaalaas: On the source-identification method. J. Acoust. Soc. Am. **104** (1998) 3546–3557.

[24] W. M. Hartmann: Localization of sound in rooms. J. Acoust. Soc. Am. **74** (1983) 1380–1391.

[25] B. R. Glasberg, B. C. J. Moore: Derivation of auditory filter shapes from notched-noise data. Hear. Res. **47** (1990) 103–138.

[26] D. Deutsch: Ear dominance and sequential interactions. J. Acoust. Soc. Am. **67** (1980) 220–228.

[27] E. R. Hafter, C. Carrier: Binaural interaction in low-frequency stimuli: the inability to trade time and intensity completely. J. Acoust. Soc. Am. **51** (1972) 1852–1862.

[28] M. P. Zwiers, A. J. V. Opstal, G. D. Paige: Plasticity in human sound localization induced by compressed spatial vision. Nat. Neurosci. **6** (2003) 175–181.