Localization in a Reverberant Room: Effects of Nearby Walls and Room Learning

Norbert Kopčo\textsuperscript{1,2} and Barbara G. Shinn-Cunningham\textsuperscript{1,2,3}

\textsuperscript{1}Hearing Research Center
Departments of \textsuperscript{2}Cognitive and Neural Systems and \textsuperscript{3}Biomedical Engineering
Boston University, Boston, Massachusetts 02215

Submitted to the Journal of the Acoustical Society of America:

Received:

Running Head: Localization in a Room

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Ba

Corresponding author’s address:
Department of Cognitive and Neural Systems
Boston University
677 Beacon St.
Boston, MA 02215

Tel: 617-353-8693
Fax: 617-353-7755
e-mail: kopco@bu.edu
ABSTRACT

A behavioral study was performed to examine the influence of listener location in a reverberant environment on spatial auditory perception. Localization accuracy was measured for sources in the right frontal quadrant of the listener's horizontal plane, with source distances between 15 cm and 1 m from the head. Listeners were positioned at four different locations in the room: in the center, with their back close to a wall, with their left ear close to a wall, or in a corner. Localization bias and variability in perceived angle from interaural axis and perceived distance were analyzed. Small but statistically significant changes in the azimuthal bias were observed as a function of listener location. The bias was smallest in the center of the room and was slightly larger when listeners were positioned with one ear close to a wall. On the other hand, a wall behind the listener caused significant increases in azimuthal bias. Listener location did not influence bias in perceived distance, even though, in general, distance perception depends on reverberation. Consistent with expectations, variability in perceived azimuth and distance increases for listener locations in which early, strong echoes are present; the influence of room location on localization performance appears to depend on the amount of distortion that the echoes and reverberation cause in localization cues. In addition to an effect of listener position, a strong room-learning effect is observed, leading to improved localization performance.
1.1 Introduction

Most of the traditional behavioral studies of auditory localization were performed in an anechoic chamber, and measured performance in only two dimensions: the perceived azimuth and elevation (Wightman and Kistler, 1989; Makous and Middlebrooks, 1990; Wenzel et al., 1993). More recently, the focus in auditory localization studies has shifted to more realistic, reverberant environments (Hartmann, 1983; Rakerd and Hartmann, 1985, 1986; Wagenaars, 1990), and the perceived source distance started to be studied as an additional experimental dimension (Bronkhorst and Houtgast, 1999; Santarelli, 2000; Zahorik, 2000b). One of the main results of these recent studies is that, compared to anechoic environment, reverberant spaces provide additional cues for distance perception, thus making it more accurate. However, reverberation also causes small degradations in directional localization accuracy (Santarelli, 2000). These degradations can be overcome by practice (Shinn-Cunningham, 2000).

In a recent study, Brown (2001) performed an analysis of the effects of reverberation on acoustic characteristics of the perceived sounds. This study showed that reverberation alters the perceived monaural spectrum of the sound, as well as the perceived interaural level and phase differences. Also, it showed that the influence of reverberation on the perceived sounds changes depending on the source position relative to the listener; and that the effect of reverberation varies as a function of the listener position in the room.
In the present study, auditory localization was examined as a function of the listener position in the room. The main parameter varied was the position and orientation of the listeners with respect to the walls of the room. Other parameters analyzed were the sound source position with respect to the listener, and the effect of experience on the performance. The analysis of the latter parameter was motivated by reports from previous localization studies in reverberant rooms (Shinn-Cunningham, 2000) where a strong room-learning effect was observed. The results of the present study were evaluated in terms of the error in left/right and distance dimensions.

Based on previous behavioral results and the acoustical analysis of reverberant rooms it was expected that for conditions near the wall, where there are strong early reflections, directional performance would be significantly deteriorated. The distance perception was expected to be influenced even more because, according to current hypotheses, the auditory computation of distance is dependent on the pattern of reverberation.

Also, observation of a learning effect was expected, manifested by decreases in subjects’ response bias and variability over time. To be able to separate this effect from the effect of the listener position in the room in the current study, the subjects were divided into two groups, and the order of tested room positions was counterbalanced between the groups.
1.2 Methods

1.2.1 Listeners

Six paid graduate students, three male and three female, participated in the study. Their ages ranged from 23 – 28 years. One subject had prior experience from participation in auditory localization experiments. All six subjects had normal hearing as determined by an audiometric screening.

1.2.2 Stimuli and apparatus

Stimuli consisted of five 150-ms-long pink-noise bursts, with 30-ms gaps between the bursts. One of five random tokens of the stimulus was chosen for presentation in each trial. The stimuli had a frequency spectrum with a 6-dB/octave roll-off from 200 Hz to 15 kHz and a 120-dB/decade roll-off out of band.

A point source developed by Brungart (1998) was used to present the stimuli. The stimuli were corrected for the non-flat spectral response of the point source. To eliminate the overall loudness as a distance cue, the level of the stimulus was crudely normalized during each presentation. In addition to this normalization, the stimulus level was roved by ±7.5 dB. With the rove, the stimulus level at the closer ear on any trial was random and uniformly distributed between 44 dBA and 59 dBA.

The pre-generated stimuli files were stored on the hard disk of the control computer. In each trial, one of the samples was randomly chosen, appropriately rescaled, and sent to the point source via a Cirrus CS4236 16-bit stereo sound card and a Crown power amplifier. Polhemus electromagnetic tracker was used to record the
Figure 1 a) Listener positions in room. Order of positions for the two subject groups (complementary-factors and conflicting-factors) shown by numerals and distinguished by font type (normal vs. outlines). b) Source locations relative to listener.

position of the subject’s head in each experimental session, as well as the actual sound source position and the subject’s response in each trial.
1.2.3 Procedure

The experiment was performed in an empty classroom (4 x 6 meters) within the Boston University Cognitive and Neural Systems Department (reverberation time of the room, $T_{60} = 0.4$ s). In four separate conditions, the subjects were positioned either in the center of the room, with their back near to a wall, their left ear near to a wall, or in the corner of the room (Figure 1a). The subjects were divided into two groups depending on the order in which they were tested at different positions. The conflicting-factors group started in the center and ended in the corner of the room, i.e., the subjects from this group started in the acoustically simplest/easiest condition, and ended in the most complex/most difficult one. It was expected that long-term exposure to the room would lead to improvement in performance over time (learning of the room acoustics), so the room effect and the learning effect were expected to act against each other for this, conflicting-factors subject group (filled numbers in Fig 1a). The complementary-factors group started in the corner and ended in the center of the room, so the effect of learning and the room effect could complement each other (outlined numbers in Fig 1a). Each subject performed four two-hour sessions. In each session, a full set of 300 trials for one room position was completed, with breaks after every 50 trials. A practice session of 50 trials proceeded the first session. Each trial started by the subject closing his/her eyes, after which the experimenter placed the source at a random, computer-proposed position around the listener, presented the stimulus, and moved the source away to a neutral position. Then, the subject was allowed to open his/her eyes and pointed to the perceived location of the sound using a wand. The sound source positions were distributed
uniformly in azimuth in one of three bins (around 0, 45, or 90°) shown in Figure 1b. Distance distribution of sources was logarithmic.

1.2.4 Analysis

The analysis of the results focused on two parameters, the perceived azimuth and perceived distance. Performance in these two dimensions was evaluated in terms of the bias (mean signed difference) and variance in responses compared to the actual positions of the sources.

1.3 Results

Initial analysis of the results showed that the improvement in performance over time (learning) is in its amount comparable to the effect of listener position in the room, which was the main effect of interest in the present study. To address this fact, the presentation and analysis of the results is divided into three parts. In Section 1.3.1, a gross analysis of the interaction between the effect of listener position and learning is presented. The analysis continues with a detailed discussion of the effect of learning (Section 1.3.2), so that the learning effect can be factored out from the analysis of the room position effect, presented in Section 1.3.3.

1.3.1 Interaction of listener position and learning

Figure 2 summarizes the overall effects observed in the experiment. The figure compares performance in the initial vs. the final experimental session, leaving out the intermediate sessions (i.e., the Back-to-wall and Ear-to-wall room positions). For the two subject groups (thick lines in Figure 2) and for individual subjects (thin lines) it shows the performance in the four analyzed parameters (mean and variance in perceived
Figure 2 Average bias and variability in subject response errors (averaged over source position) in the initial and the final experimental session (Session 1 indicates the Center listener position for the balanced-factors group and the Corner listener position for the unbalanced-factors group. In session 4 the groups are reversed). a) left/right response bias; b) left/right variance; c) distance response bias; d) diastase response variance.
azimuth and distance) as a function of the session number. The conflicting-factor group started (session #1) in the center of the room and ended (session #4) in the corner. For the complementary-factor group the room positions were reversed.

Figure 2a shows the change in the azimuthal bias between the two sessions. The conflicting-factors group changed from essentially no bias at the beginning to slight medial bias at the end of experiment. The complementary-factors group was strongly biased to respond closer to the median plane at the beginning, and this bias has decreased over time. So there was no consistent change in the azimuthal bias over time. There are two possible explanations for the differences observed in Figure 2a. First, it is possible that there is no temporal effect on this parameter and the observed differences (the magnitude which was approximately 2.5° for both groups) are solely due to the change of the room position. That is, the result is consistent with the statement that when a listener moves from a center of a room to a corner, his/her responses become more biased towards the median plane by 2.5° (one-tailed $t$-test on average bias for six subjects and six source bins from Figure 2b, $t_{35}=2.109$, $p=0.021$). Alternatively, it is possible that only the magnitude of the bias was decreasing over time, independent of its initial direction. This decrease is on average 1.4° ($t_{35}=2.135$, $p=0.020$). Although it is not possible to directly decide which alternative is correct based solely on the presented data, the detailed analysis in Section 1.3.3 shows that the effect is a result of a change in the room position.
Figure 2c shows the bias in the perceived distance for the two subject groups. The largest difference observed is the difference between the two groups, caused by a strong and consistent tendency of two of the subjects from the complementary-factors group to overestimate distance (on average, the conflicting-factors group overestimated distance by 3.6%, whereas the complementary-factors group by 20.2%). There is also a small consistent trend to increase the amount of overestimation over time. On average, subjects overestimated distance by 9.3% in the initial session and by 15.2% in the final
Figure 3 Histograms of change in performance over time (Session 1 vs. Session 4) for the balanced-factors group (dashed lines) and the unbalanced-factors group (full lines). Separate histograms for overall performance (left-hand graph), performance for near sources (center graph), and far sources (right-hand graph). a) difference in azimuthal response biases; b) ratio of azimuthal variances; c) ratio of distance response biases; d) ratio of diastase response variances.
session ($t_{35}=2.245$, $p=0.015$). On the other hand, the room position doesn’t influence the perceived distance at all.

The variance in the perceived azimuth and distance is shown in Figure 2b and Figure 2d, respectively. These graphs show an interaction between the temporal effects and the effect of the room position. For the complementary-factors group (full lines) the variability always decreases over time, whereas for the conflicting-factors group (dashed lines) the variability is essentially constant. This result suggests that listeners are getting better (more consistent) in their responses over time, and that their responses are less variable in the center than in the corner of the room. The latter observation can be explained by the acoustical properties of the perceived sound. In the center of the room, there are no near walls and the perceived reverberation is essentially independent of the sound source position. Also the amount of reverberation is relatively low here. In the corner, the pattern of reverberation is dominated by the early reflections of the nearby walls, the magnitude of which can be comparable to the magnitude of the direct sound. Since the magnitude and the delay of the early reflections are changing dramatically as a function of the source position, it is expectable that these reflections will cause an increase in the variance of responses.

1.3.2 Effect of learning for near vs. far sources

Figure 3 presents an analysis of the interaction between the effect of learning and the listener’s position in the room, as a function of the sound source distance. The four parameters analyzed are, again, the mean signed error of the response and the response variance, for both perceived azimuth and distance. In each panel, the data for the two
subject groups are shown as histograms of the change in the given parameter from
session 1 to session 4 (Figure 3a shows a difference, the other graphs show a ratio of
values for session 4 vs. session 1). Three separate histogram pairs are shown in each
panel, one for the overall performance (left panel), one for near (center panel) and one for
far (right panel) sources.

The left-most graphs in each panel show the distribution of the 36 points
(6 subjects x 6 source position bins) that were averaged to generate plots in Figure 2.
Because of the quantization necessary to produce the histograms, some trends obvious in
Figure 2 are not so clear. For example, the dashed line in the left panel of Figure 3a peaks
at 0, although the mean value from Figure 2a is –2.5°. The two figures showing the
amount of bias (Figure 3a for azimuthal bias and Figure 3c for distance bias) do not show
clear trends as a function of the source distance. Figure 3b and Figure 3d show the
histograms of change in response variability in perceived azimuth and distance,
respectively. As shown in the previous section, the overall graphs for the
complementary-factors group show a strong decrease in the response variance over time,
whereas for the conflicting-factors group the two factors, learning and room position,
cancel each other and the trends are very weak. However, separate plots for near (center
graphs in Figure 3b and Figure 3d) and far sources (right-most graphs) show that the
effect is dependent on the source distance. Specifically, all the histograms for far sources
(right-most graphs) show a decrease in variance over time for both subject groups and
both parameters. For near sources, the complementary-factors group still shows a
decrease in variance over time, whereas the conflicting-factors group shows either no
change (dashed line in center graph of Figure 3b) or a clear increase in variance (dashed line in the center graph of Figure 3d). In summary, the temporal (learning) effects are the main factor causing the change in the amount of response variance for far sources (especially for the azimuthal response, Figure 3a), whereas the room position effect is equally important (Figure 3a) or dominates (Figure 3?) the responses for the near sources.

1.3.3 Detailed analysis and discussion

Figure 4 presents a detailed analysis of the results of this behavioral study, with focus on the effect of the room position. Presented are the same statistics as in the previous figures (mean and variance in perceived azimuth and distance), averaged across all the subjects (thick line), and separately across the two subject groups (thin lines), to show the extent to which the data are influenced by learning effects. Data are plotted as a function of room position (Center, Back to wall, Ear to wall, and Corner), with a separate graph for each of the six source position bins shown in Figure 1b. Figure 4a shows the results averaged across the source position bins. Figure 4b shows the data separately for each source bin.

1.3.3.1 Mean azimuthal error

The first row of panels in Figure 4 shows the mean difference between the actual and perceived azimuth of the sound sources. The plot of results averaged across the source positions (top panel in Figure 4a) shows that there is no bias at all (mean bias is 0.4° towards median plane) when the subjects are in the center of the room (left-most point). On the other hand, as discussed in Section 1.3.1, there is a significant amount of
bias towards the median plane (mean 3°) when the subjects are placed in the corner of the room (Figure 4a, right-most point). This effect of room position is even stronger when the subject is seated with his/her back against the wall (Figure 4a, second point from the left).

The bias towards the median plane is in this case on average 3.8° (significance of difference re. Center position, $t_{35}=3.803$, $p=0.00027$). Subjects seated with their left ear towards the wall do not show any significant bias (mean bias is 1.2° towards median plane). This shows that nearby walls do influence the listener’s ability to correctly localize sounds in the azimuthal dimension in a reverberant room.

The main aspect that influences the bias seems to be the presence of a wall behind the listener, which, in the present experiment, was the case in the Back-to-wall and the Corner conditions. When data are analyzed separately for each source position bin (shown in the top row of Figure 4b), this hypothesis is supported by data in three out of six source position bins (0° 15 cm, 45° 1 m, and 90° 1 m).

At some of the source positions shown in Figure 4b there are also significant differences between the two subject groups. For example, an interaction between the effect of room position and learning is observed for targets directly ahead of the listener (first and fourth graph from the left in Figure 4b). The interaction is shown in that, for both of these source position bins, the conflicting-factors group’s performance in the Back-to-wall condition is biased towards the median plane, compared to the complementary-factor group’s performance, whereas in the Ear-to-wall condition the complementary-factors group’s performance is biased more laterally (The trend would be
easier to visualize if only the differences between the two groups were plotted). A possible explanation for this observation is that the temporal (learning) effects are more complex for sources in front of the listener than for other source positions. Namely, if the data is analyzed as a function of time there is no significant bias (re average) in the first session. However, in the second session there is a strong bias towards the medial plane. In the third session the bias reverses and the subjects respond more laterally, and in the fourth session the bias effectively vanishes. This observation can be an artifact of a process that takes place when the listener enters a new room and familiarizes himself with the new acoustic environment. This familiarization process can be very slow and changes can occur over periods of hours (Shinn-Cunningham, 2000). However, it is not clear why the result is that the direction of the bias changes over time. Also, it is notable that the effect occurs only for sources ahead of the listener where the direct sound activates both ears approximately equally (although there are some differences among the subject groups also for other source positions). It is in general assumed that the acoustic characteristics of perceived sounds that change from room to room are mostly the monaural characteristics (amount of reverberation, spectral shape). Then, these are the characteristics that the listener needs to adapt to repeatedly for every room. And the fact that the adaptation to individual room positions is most visible for the source positions in front of the listener suggests that having approximately the same amount of direct sound at both ears is important for this adaptation process.
Figure 4 Localization performance as a function of the listener position in the room (Center of the room, Back to wall, Left ear to wall, and Corner). A) Average across source position bins. B) Individual columns show data for different source position bins (combinations of 0, 45, and 90° by 15 cm or 1 m). Rows show the four statistics of azimuthal bias, azimuthal variance, distance bias, and distance variance (top to bottom, respectively). Overall average is shown by the thick line, group means are shown by the thin lines.
1.3.3.2 Variance in azimuthal perception

The second row of graphs in Figure 4a and Figure 4b shows the variance in perceived azimuth as a function of room position. Results averaged across source positions are shown in Figure 4a. Averaged across subject groups (thick line), the results are in accord with expectations. I.e., the amount of variability in subjects’ responses grows as they are moved from the center to the corner of the room, that is, from the simplest to the most complex acoustic environment. This result is to be expected because of the number of early strong reflections that influence the perceived sound. This number is very high when the listener is in the corner of the room and the structure of these reflections changes dramatically with small changes in the sound source location, thus leading to increased variance in the perceived location. Also the gradual growth in the amount of variation for the middle positions (back-to-wall and ear-to-wall) is consistent with this hypothesis. Comparing the two subject groups (thin lines) shows a very strong temporal effect on the amount of variance in the perceived azimuth. Especially, the conflicting-factors group actually improves when moved from the center to the corner of the room despite the increase in the complexity of the acoustic environment. This result means that, when looking at the variance in the perceived azimuth, the influence of experience is stronger than the influence of the room position.

The second row in Figure 4b shows the amount of variance in azimuthal perception, binned by the sound source positions with respect to the listener. One clear difference is in the amount of variance between the near bins (the three left-most graphs)
and far bins (the three right-most graphs), where the variance is much larger for the near source bins. However, this effect can be unrelated to the interaction of the room position and the source position. Instead, it can be an artifact related to the amount of angular error in the pointing operation that the subject performs to indicate the perceived source position, because this amount is larger for the near than for the far sources. Another trend visible in the data in the second row of Figure 4b is that the amount of variance in perceived azimuth increases with the source azimuth (i.e., with the angle from the median plane). This trend is present for both near and far sources, except for the near sources in the ear-to-wall room position.

1.3.3.3 Mean distance error

The third row of graphs in Figure 4 shows the mean error in the perceived source distance. On average, the subjects tend to overestimate the source distance. This effect is almost completely due to the two subjects in the unbalanced-factors group who constantly overestimated the distance, as discussed in Section 1.3.1. The average graph (Figure 4a) shows that the perceived distance is almost completely independent of the position of the listener in the room. This result contradicts most current theories of distance perception (Bronkhorst and Houtgast, 1999; Santarelli, 2000). These theories assume that the computation of perceived distance involves estimation of the amount of energy in the part of the sound that arrived at the ears directly from the sound source (direct part) and the amount of energy in the portion that was reflected off the walls (reverberant part). The amount of reverberant energy changes dramatically as the listener’s position in the room changes (Brown, 2001). Also, this amount is more or less
independent of the azimuthal position of the sound source when the listener is in the center of the room, but it changes with azimuth considerably when the listener is near a wall. However, neither of these two factors seems to influence the data shown in the graphs in Figure 4. Assuming that the hypothesis about the involvement of reverberation in the brain’s computation of perceived distance is correct, there are two possible explanations of this result. Either the parameter derived from the amount of reverberation is independent of the room-position-related changes in the reverberation pattern, or the brain is able to compensate for these changes automatically.

The results binned by the sound source position are shown in the third row of Figure 4b. Again, the results averaged across subject groups show essentially no effect of room position. However, there is a clear influence of the sound source position on the perceived distance. For far sources (distances between 38 cm and 1 m) there is essentially no bias. For near sources (distance less than 38 cm) there is a strong bias to overestimate distance. This bias is strongest for sources in front of the listener, and it essentially vanishes for sources on the interaural axis. The trend to overestimate distance for near sources is opposite to the results for sources farther than a meter from the listener where the trend is to underestimate distance (Zahorik, 2000b).

Comparison of performance between the two subject groups (thin lines in the third row of Figure 4b) shows that, although there is no effect of room position, there are some temporal changes in the perceived distance. The two thin lines in all graphs for azimuths of 0° and 45° (first, second, fourth, and fifth graph from the left in third row of Fig 4b) are approaching each other as the room position changes from Center to Corner.
This means that the tendency to overestimate the source distance was increasing with time. However, for sources on the interaural axis (third and sixth graph) there is either no trend (sixth graph) or the temporal effect actually reverses (third graph).

When comparing results for the average bias in azimuth and distance (first vs. third panel from top in Figure 4a) one can notice that there is no bias in either of the two parameters for the balanced-factors group in the center of the room. This could suggest a hypothesis that when a listener enters a new room, the process of adaptation to (or estimation of) the acoustic parameters of the room assumes, or requires, that the listener is in the center of the room. If this condition is not fulfilled, whether because the listener was not in the center of the room when he/she first entered the room (complementary-factors group) or because he/she is not in the center any more (conflicting-factors group), the performance in azimuth or distance is biased.

1.3.3.4 Variance in distance perception

The bottom row of graphs in Figure 4 shows the results in terms of the variance in perceived distance. The data averaged across source positions are shown in Figure 4a. Similar to the results for variance in the azimuthal perception, the variance in perceived distance is smallest when the listener is in the center of the room, and largest when he/she is in the corner. However, the intermediate room positions (Back to wall, Ear to wall) are this time almost identical with the center of the room position, which means that the complexity of acoustic environment has to increase dramatically to influence the consistency in perceived distance judgments. Comparison of performance between the two subject groups shows that the effect of learning in this parameter was the smallest of
all the four parameters analyzed, opposite to the conclusion drawn in Section 1.3.1 based on Figure 2d which ignored the intermediate positions.

When the data are binned by the sound source position (bottom row of plots in Figure 4b), no effect of room position or learning is visible for far sources (three right-most plots). The increased variance in distance perception for the Corner of room position holds only for near sources (three left-most plots). As was the case with the variance in perceived azimuth, this effect could be an artifact related to the inaccuracy in the listener’s pointing to perceived position, which is larger for near than for far sources. However, the fact that the error is independent of distance for three of the four room positions suggests that the increased variance in the corner of the room is an effect caused by the room acoustics. The only point where there is a large difference in performance between the two subject groups is the Corner point in the 90° 15 cm graph, which supports the conclusion that the temporal effects are negligible for the variance in perceived distance.

1.3.4 Summary

The data presented in this chapter show that both nearby walls and experience with a room influence listener’s ability to accurately localize sounds. Moreover, a complex interaction between the two factors was shown. The position of a listener in a room influences perception in terms of azimuthal bias, as well as the variance in perceived azimuth and distance. The learning effect dominates performance in terms of variance in perceived azimuth and, to a lesser extent, in terms of bias and variance in perceived distance. The effects are stronger for nearby sources (up to 38 cm) than for far
sources (38 cm to 1 m) in all four parameters except for the azimuthal bias, where the effects are comparable.

1.4 **Comparison of behavioral results and acoustic measurements in the center of the room**

The results summarized in the previous section show that there is a complex relationship between the acoustic characteristics of different listener positions in a room and the localization performance at these positions. The goal of this section is to gain insight into how the pattern of reverberation, i.e., the main room-position-dependent acoustic perceptual factor, influences the cues for auditory localization, and thereby the localization performance. For this purpose, an analysis is performed on a set of Head-Related Transfer Functions (HRTFs) measured as a function of the sound source position with respect to the listener, with listeners positioned in the center of the experimental room. This narrows down the focus from the previous sections where the changes in performance were studied as a function of both the listener position and the sound source position. The changes in acoustical cues as a function of the source position are stronger and less complex than the ones connected with the changes in the listener position. Because of that the effect of the source position changes on behavior should be stronger and more reliable.

The effect of reverberation can be both beneficial and detrimental. For example, Santarelli (2000) showed that addition of reverberation improves localization accuracy in the distance dimension, while worsening performance in the azimuthal dimension. This section is studying four acoustic parameters important for auditory localization. These
are the near- and far-ear monaural spectra and the interaural level and time differences. The analysis looks at the variability in these parameters as a function of frequency evaluating two hypotheses about the effect of the variability. First, if any of the parameters is important for auditory localization in some dimension then increased variability in that parameter should imply worsening of the performance in that dimension. Second, the brain itself might be able to compute the variability in these parameters, and use it as a new parameter (i.e., new localization cue), thus improving the localization performance in reverberant compared to the anechoic environment.

The candidate parameters that could well characterize the effect of reverberation on auditory localization should have several properties. First, their value should be constant in the anechoic environments and change only in the reverberant ones. This property would mean that the parameters really measure only some characteristic of the reverberation and thus explain how reverberation influences performance. Also, they should ideally not require extraction of the HRTF from the perceived sound because that is something that the brain is not capable of when it computes the sound location. However, this would be too strict a requirement, so here we loosen it a little bit by requiring only that the parameters do not assume the brain to be able to separate the direct-sound part of the HRTF from the reverberant part. Specifically for the binaural parameters, this requirement essentially makes the parameters computable even without availability of the HRTF, as long as the presented sounds are sufficiently broadband.

The chapter is divided into the following sections. Section 1.4.1 summarizes the methods used to measure and analyze the HRTFs. Section 1.4.2 presents an analysis of
the variations in acoustic cues caused by reverberation. Several candidate parameters are suggested that can explain the effect of reverberation on localization performance. In Section 1.4.3, these candidates are compared to the center-of-room data from the present behavioral study.

1.4.1 Methods and analysis

The HRTFs discussed in this section were collected and first analyzed by Brown (2001). Brown measured the HRTFs using fifteen human subjects and the KEMAR acoustic manikin. In her study, the listeners (and KEMAR) were positioned at one of four locations in the same reverberant room as was used for the present behavioral study (Section 1.2). The analysis in this section is focused on data obtained with listeners in the center of the room. The HRTFs were measured for six sound source positions around the listener. The six positions were combinations of the azimuth of 0, 45, and 90° and distance of 15 cm and 1 m, all in the right frontal quadrant of the listener’s horizontal plane. The Maximum-Length Sequences (Zahorik, 2000a) were used as stimuli. In every presentation, the stimuli, stored on a PC computer, were converted into analog signals using the D/A converter of the TDT PD1 system, amplified by a Crown amplifier, and presented using the BOSE cube speaker. The responses were recorded at the entrance of the listener’s ear canal by Knowles Electret microphones mounted onto ear-canal-blocking earplugs. The signal recorded by the microphones was converted by the TDT PD1 A/D converter and stored onto the PC’s hard-drive. Responses to 10 stimulus presentations were averaged before the resulting impulse response (i.e., the HRTF) was computed.
The impulse responses were 750 ms long and were sampled at 44.1 kHz. This contained all the reverberation of the room, as well as the direct-sound HRTF. To separate the effect of reverberation on the localization cues, a pseudo-anechoic HRTF was derived from each reverberant HRTF by windowing out the reverberant part. A time point of 5.5 ms after the on-set of the direct-sound portion of the HRTF was chosen for the windowing. At this time point the direct sound has essentially died out while none of the reflections has yet arrived at the ear.

Four reverberation-dependent parameters were evaluated in this study: variability in the near-ear magnitude spectrum, variability in the far-ear magnitude spectrum, variability in the interaural level difference, and variability in the interaural time difference. For all of them, the processing was as follows. First, every HRTF was zero-padded to the length of 44,100 taps, then transformed by the discrete-fourier-transform (DFT). For the magnitude-related parameters (variability in the monaural spectra and ILD), the third step was to convert the magnitudes into the dB units and smooth them by 1/12-octave smoothing filter. Let $s_L$ and $s_R$ be the resulting left and right magnitude spectrum, respectively. Then the variability in the monaural spectra was computed as

$$V_L = 1000 \ \text{mean}\left\{\frac{s_L(201)-s_L(200)}{200}, \frac{s_L(202)-s_L(201)}{201}, \ldots, \frac{s_L(12001)-s_L(12000)}{12000}\right\}$$

i.e., as the mean absolute frequency-weighted frequency-to-frequency difference in the level in the range of 200 to 12000 Hz. The computation of variability in the ILD was
similar, except that it was performed on the difference between $s_L$ and $s_R$ at any given frequency. When computing the variability in the ITD, the third step was to obtain the difference between the left and right phase spectrum and to find a correct number of periods to shift the spectrum to the appropriate interaural time difference. To obtain the period shift it was also necessary to derive the difference in the phase spectra of the corresponding pseudo-anechoic HRTFs, and unwrap this phase spectrum difference, because was much smoother and much better behaved than the reverberant one. Then, the period shift that minimized the difference between the reverberant and anechoic phase spectrum was used at each frequency in the reverberant phase difference spectrum. The phase-difference spectrum obtained in this way was then smoothed by 1/12-octave smoothing filter and the ITD-variability parameter was computed on the resulting vector, $\phi$, as

$$V_{ITD} = 2\pi \ \text{mean} \left\{ \frac{\phi(201) - \phi(200)}{200}, \frac{\phi(202) - \phi(201)}{201}, \ldots, \frac{\phi(12001) - \phi(12000)}{12000} \right\}$$

i.e., as the mean absolute frequency-to-frequency difference in interaural time differences in the range of 200 to 12000 Hz.

1.4.2 Acoustic measurements

Figure 5 compares the anechoic and reverberant magnitude spectra at the six sound source positions. The figure shows that the main effect of reverberation is the addition of frequency-to-frequency variability to the monaural spectra. This variability is
most visible at high frequencies. Another effect is that the reverberation fills in high-frequency notches in the far-ear spectra. Both these effects also influence the interaural level differences (ILD) so that in the reverberant HRTFs there is increased frequency-to-frequency variability in the ILD and decreased high frequency ILDs.

Figure 6 shows the anechoic and reverberant interaural time differences (ITD) at the same six sound source locations. Similar to the above amplitude-related parameters, the main effect of reverberation is in increased frequency-to-frequency variability in the ITDs. However, now the increase is most significant at the low frequencies.

In both Figure 5 and Figure 6 the main effect of reverberation is in the increased frequency-to-frequency variance. It can be expected that this increase should lead to decrease in the accuracy of judgements of the perceived source location. However, this effect is source-position dependent, and so if the brain can compute the amount of variance, the amount can be used as another auditory localization cue, leading to improved performance. To test these hypotheses, Figure 7 shows the amount frequency-to-frequency variability in all four of proposed parameters as a function of the sound source position. Plotted is variability, computed as described in Section 1.4.1, for each of the 15 human subjects and KEMAR and in both anechoic and reverberant environment. Comparison of Figure 7a vs. Figure 7b shows that reverberation adds variability to all four studied parameters. The amount of variability added by reverberation (Figure 7b) is always sound source position dependent. The variability in
Figure 5 Anechoic and reverberant magnitude spectra measured on KEMAR acoustic manikin positioned in the center of the room. The higher of the curves in each panel corresponds to the ear closer to the sound source (right ear in the present measurements), the lower is the far (left) ear. Sound source was positioned at six different locations with respect to KEMAR.
Figure 6 Interaural time difference in the anechoic and reverberant HRTFs measured with KEMAR in the center of the room. The full lines are the ITDs in the anechoic responses and the dots are the reverberant ITDs. Individual panels correspond to sound source positioned at six different locations with respect to KEMAR.
Figure 7 Frequency-to-frequency variability in (top to bottom rows) left-ear (far-ear) monaural spectrum, right-ear (near-ear) monaural spectrum, ILD, and ITD, as a function of sound source position for KEMAR and 15 human listeners. a) Anechoic HRTFs. b) Reverberant HRTFs.
these parameters in the anechoic environment (Figure 7a) is relatively source-position independent, and for most of them (except for the ITD-variability), the amount of variability in the reverberant environment is always larger than in the anechoic space. So, all these parameters can be used as localization cues. The effect of reverberation (Figure 7b) on the studied parameters is consistent as a function of the sound source position. Variability is always larger for far sources than for near sources. For the far-ear monaural spectrum (top panel, Figure 7b) and the binaural parameters (the bottom two panels, Figure 7b) the variability increases with source laterality. On the other hand, for the near-ear spectrum (second panel from top, Figure 7b), variability decreases with source laterality.

Several hypotheses can be proposed based on the above observations. First, the left/right response variance may increase with source laterality and distance, due to the increased variability in the low-frequency ITD and high-frequency ILD (two bottom panels in Figure 7b) which are known (Duplex theory) to contribute to auditory localization. The cues for perceived distance are much less well understood. In anechoic spaces, ILD can be used (Brungart, 1998). In reverberant spaces, reverberation contributes to distance judgements (Santarelli, 2000). If similar trends are observed for one of the parameters in the variability plots in Figure 7 and in behavioral data, this might suggest that humans use the parameter to judge distance of auditory objects.
1.4.3 Behavioral measurements

To evaluate the hypotheses proposed in the previous section, Figure 8 shows the localization data collected in the present behavioral study (Section 1.3) with listeners in the center of the room. Individual data for each listener are shown.

Figure 8a shows the bias in the listeners’ azimuthal perception. There are no clear trends in performance as a function of the sound source position. This is especially the case if the differences in performance among subjects are considered.

The variance in azimuthal perception is shown in Figure 8b. As predicted, the variance grows with the source laterality. Part of this effect can be due to the variation in the binaural cues caused by the reverberation. But this trend can be also due to psychoacoustics factors, namely the large azimuthal errors near the interaural axis can be due to smaller changes in binaural parameters as a function of source position in this region. The left/right variance also decreases with distance. This trend is opposite to what was predicted in the above analysis, which suggests that the acoustical factors do not dominate the performance. Again, large inter-subject differences can be observed in the data, which suggests that the performance is not dominated by the reverberation-related acoustical factors.

The bias in distance perception is shown in Figure 8c. The bias decreases with the source distance. Subjects overestimate distance of near sources, but are accurate for source distances near 1 m. Also, bias depends on source laterality for near sources, but not for the far sources. Neither of these trends correlates with the acoustical analysis
Figure 8 Mean and variance in localization judgments as a function of source position. Each bar represents one listener. a) left/right bias, b) left/right variance, c) distance bias, d) distance variance.

presented above. On the other hand, previous studies (Santarelli, 2000) have shown that the distance perception in reverberant environments is significantly improved compared to anechoic space. And this finding can be supported by the trends in binaural variability cues in Figure 7.

Figure 8d shows the variance in the distance judgements. The variance decreases with the source laterality for nearby sources, as well as with source distance. Again, this
trend is opposite to the changes in the amount of variability in localization cues proposed by the acoustic analysis.

1.4.4 Summary

There is a complex interaction between the effect of reverberation on the acoustic localization cues and the localization performance. The acoustic analysis showed clear and consistent trends in the effect of reverberation on the localization cues. But behavioral measurements show that these trends contribute only partially to the observed performance and that they can be easily outweighed by other factors like the inter-subject differences. The results suggest that the amount of frequency-to-frequency variability is inversely proportional to the variance in performance in the azimuthal dimension and the distance dimension, as a function of the source distance. As a function of source azimuth, the variance in perception in the azimuthal dimension can be assigned to the variability in the far-ear monaural spectrum or any of the binaural parameters. The variance in the perceived distance is then more correlated with the variability in the near-ear monaural spectrum.

1.5 Conclusions

The results of the behavioral experiment presented in this chapter show that the position of a listener in a room has a strong impact on his/her localization behavior. It was also shown that the ability to accurately localize sounds is influenced by the listener’s familiarity with the room. The variance in response in both the azimuth and distance dimensions is larger in the corner than in the center of the room, i.e., it tends to be large when a listener is near a wall, but decreases with experience. Also, the bias in
perceived distance is not affected by room position or experience, despite the previously shown importance of reverberation for distance perception. In general, localization accuracy is influenced by nearby walls more for near sources than for far sources.

The analysis of acoustic measurements has shown that reverberation causes systematic degradations in acoustic localization cues. The amount of these degradations varies with the level of the direct sound. The reverberation-related variations were shown to be largest for far sources and their amount was shown to change with the source laterality. When trying to relate these results to behavioral data, the acoustic effects were confounded with other psychoacoustic factors, thus making strong conclusions problematic. Despite that the results support the hypothesis that laterality is determined by the binaural cues and that in reverberant rooms near-ear monaural cues contribute significantly to the judgements about perceived distance.
REFERENCES


Brungart DS (1994) Distance information transmission using first order reflections. In.: Massachusetts Institute of Technology.


