ACOUSTICS AND PERCEPTION OF SOUND IN EVERYDAY ENVIRONMENTS

Barbara Shinn-Cunningham

Boston University
677 Beacon St.
Boston, MA 02215
shinn@cns.bu.edu

ABSTRACT

One aspect of hearing that has received relatively little attention by traditional psychophysicists is how echoes and reverberation in everyday spaces affect perception. In the ordinary world, echoes and reverberation are ubiquitous and influence the signals reaching the listener, the processing of these signals by the brain, and the resulting perception of both sound sources and the environment. Many aspects of the signals reaching the ear are altered or "distorted" by echoes and reverberation, including spectral content, interaural differences, and temporal structure. As a result, echoes and reverberation could influence many aspects of perception, including spatial hearing in direction and distance, speech intelligibility, and spatial unmasking. This paper reviews results from a number of studies examining how the acoustics of ordinary rooms affect various aspects of the signals reaching a listener's ears as well as resulting perception. While the acoustic effects of reverberant energy are often pronounced, performance on most behavioral tasks is relatively robust to these effects. These perceptual results suggest that listeners may not simply be adept at ignoring the signal distortion caused by ordinary room acoustics, but may be adapted to deal with its presence. These results are important for designing truly immersive spatial auditory displays, because they demonstrate the importance of reverberant energy for achieving a realistic, immersive experience.

1. OUR COMPLEX ACOUSTIC WORLD

Most psychology and neuroscience textbooks discuss auditory perception as if the only acoustic energy reaching a listener arrives directly from a source, ignoring the fact that much of the energy arrives indirectly, reflecting off the many objects in the environment. In fact, in decoding sound, the auditory system deals with a much more complex and interesting set of problems than the simplified textbook view suggests, and does so efficiently and elegantly.

For instance, textbooks treat the computation of sound source location as a relatively straightforward problem, as if the auditory system simply extracts basic acoustic cues (such as interaural differences and spectral cues) and estimates the location of a source from these cues (e.g., see [1, 2]). This analysis assumes that acoustic cues that arise for a source from a particular location do not vary with the environment and virtually ignores the computation of source distance. However, in our everyday lives, the very cues used to compute source position depend not only on the sound that reaches the ears of the listener directly, but on sound that is reflected off of all of the objects in the environment, including walls, floor, furniture, etc. As a result, the acoustic cues used to determine source position depend not only on the location of the source relative to the listener, but also on the acoustic environment and the location and orientation of source and listener in the environment [3, 4]. It is well known that the presence of reverberant energy provides information about acoustic source distance [5-15], but little is known about listeners compute source distance based on the signals reaching their ears [15, 16].

Similarly, most discussions of speech recognition focus on how to decode the signal emitted by a talker, not the actual signal reaching the receiver (e.g., see [17]). Most automatic speech recognition systems are designed to interpret clean speech (without any reverberation) and therefore tend to perform poorly in ordinary reverberant environments. In contrast, the ability of human listeners to understand speech is relatively robust in the presence of modest amounts of reverberant energy [18-20].

This paper reviews a number of studies investigating how reverberation in a moderate-sized room influences acoustics and perception. Results suggest that the presence of echoes and reverberation (referred to jointly as "reverberation" throughout the remainder of this paper) significantly distorts many of the acoustic cues thought important for spatial hearing and other perceptual tasks. However, listeners are not only adept at making accurate judgments in ordinary reverberant environments, they are in fact adapted to the presence of reverberation and benefit from its presence in many ways. These results underscore the importance of simulating room acoustics in order to create realistic and natural three-dimensional spatial auditory displays.

The initial studies reviewed in this paper focus on studying the acoustics of a moderate-sized classroom (dimensions of 5 m x 9 m x 3.5 m). Of course, the effects of reverberation vary dramatically with environment. However, studying in depth what happens due to the reverberation present in one particular environment provides general insights into both acoustic and perceptual effects of reverberation in other relatively small rooms.

In order to understand how the effects of reverberation are influenced by listener location, these studies compare effects for four different locations of the listener in the room, depicted schematically in Figure 1. In the ***center*** location, there are no reflective surfaces near the listener. In the ***left*** location, the listener's left side is very close to one wall. In the ***back*** location, the listener's back is to the wall. Finally, in the ***corner*** location, both the listener's left side and back are near a wall. For all of these listener locations, a sound source

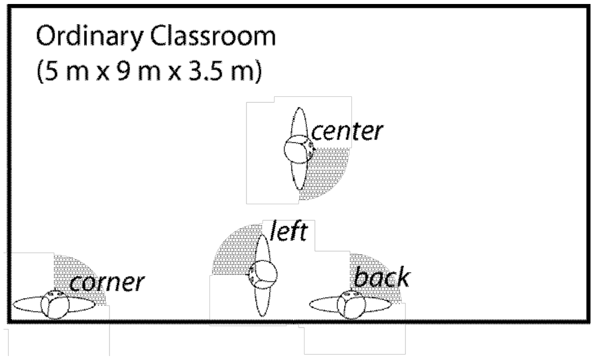

Figure 1. *Sketch of the four listener locations in the room (gray area shows source locations re: listener).*

was presented from a range of sources in the front right quadrant of space (depicted in gray in Figure 1). In some cases, acoustic measurements were used to analyze how listener location influenced the signals reaching the ears. In other cases, perceptual studies were performed in the actual room. Finally, acoustic measurements were used to create realistic headphone-based simulations of the signals reaching the listener in order to study the perceptual effects of realistic reverberation with carefully-controlled stimuli.

2. ACOUSTIC EFFECTS OF REVERBERATION

Reverberation influences nearly all acoustical attributes of the signals reaching the ears of the listener, including temporal structure, spectral content, intensity, and interaural differences. Simple physics dictates that the magnitude of all of the acoustical effects of reverberation varies inversely with the direct-to-reverberant energy ratio (D/R). The direct sound level reaching the ear varies dramatically with the position of the sound source relative to the listener; as a result, D/R changes systematically with source position [21-23]. The direct sound level decreases with the square of the distance from source to head; thus, D/R decreases and the effects of reverberation increase with distance [22]. As the sound source is moved laterally off the median plane, the direct sound level decreases at the far ear and increases at the near ear, an effect that is more pronounced at high frequencies and increases as the source approaches the head. Thus, the effects of reverberation vary with source laterality (and do so most strongly for sources near the listener), increasing at the far ear and decreasing at the near ear [22].

The location of the listener in the room also has a dramatic effect on the level and pattern of reverberation at the ears [4, 21-26]. For instance, when the listener is far from any walls, the steady-state effect of the reverberation is a distortion that is essentially independent from frequency to frequency; however, early, strong reflection (such as arise when a listener is seated close to a wall) lead to systematic distortions of acoustic spatial cues. When a listener is near a reflecting wall, the spectral content of the total signal reaching the listener is comb-filtered (due to frequency-dependent reinforcement and cancellation of the direct sound by the echo), producing notches in the spectra received at the ears that vary with listener position relative to the wall and

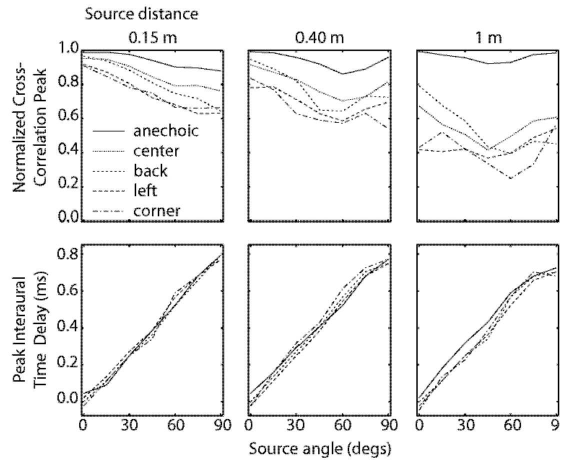

Figure 2: *Cross-correlation analysis of left and right ear signals as a function of source location and listener location. Top row shows the normalized cross-correlation peak amplitude. Bottom row shows the interaural time delay of the peak.*

source position relative to the listener. Such comb-filtering also systematically distorts interaural differences.

Whether or not there are early, intense reflections, the overall effect of reverberation is to introduce variations in the interaural and spectral localization cues that govern localization perception, even though these acoustic effects are more pronounced when a listener is near a wall and when the source is far from the listener.

Figure 2 demonstrates how reverberation influences the interaural time difference (ITD) for different listener and source positions. Head-Related Transfer Functions (HRTFs) were measured for source locations in the horizontal plane in the front-right quadrant of space relative to KEMAR for the four listener locations shown in Figure 1. The ***center*** HRTF locations were time-windowed to generate pseudo-anechoic HRTFs. The left- and right-ear HRTFs for the resulting five sets of HRTFs (***anechoic***, ***center***, ***back***, ***left***, and ***corner***) were bandpass filtered from 100 – 2000 Hz, cross-correlated, and then normalized by the energy in the bandpass-filtered left- and right-ear HRTFs. The main peak falling within the physiologically-plausible ITD range (-1 to +1 ms) was identified. The normalized peak magnitude (which has a maximum value of +1 by construct) is plotted in the top row of Figure 2; the corresponding ITD value is plotted in the bottom row of the figure.

In anechoic space, the peak correlation value decreases, albeit slightly, with source azimuth due to frequency-dependent filtering of the head. When the source is in front of the listener, the left- and right-ear signals are nearly identical; however, the similarity decreases with source azimuth. The general effect of reverberation (regardless of listener location) is to decrease the peak correlation value, an effect that grows with both source laterality and distance. The peak value also depends on the listener location in the room, and generally is smallest when the listener is in the ***corner*** or ***left*** room locations (where the head-shadowed ear receives early, intense reflections).

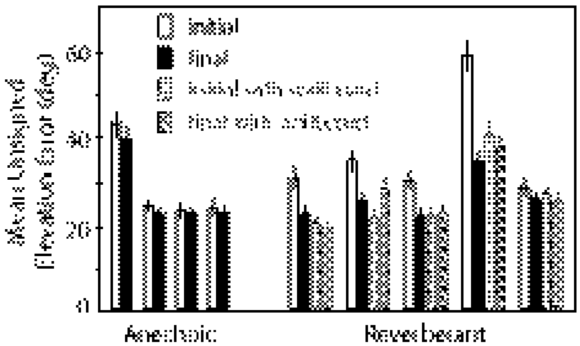

Figure 3: *Mean error in elevation (deg) for the first and final 200 trials in each condition of a localization pointing experiment. Each set of bars represents results from one subject. Error bars show standard error over the 200 trials.*

The effects of reverberation on binaural cues are dominated by the effect of reverberation on the ear with the smaller D/R. As a result, the D/R at the ear that is farther from the source determines how distorted ITD is by reverberation. This D/R decreases with source distance and azimuth and depends on listener position. In particular, when a listener has his shadowed ear facing a wall, peak correlation values are very low, especially for distant, lateral sources.

Despite the fact that the peak correlation value is dramatically affected by distance and listener location in the room, the ITD at which the peak occurs is relatively insensitive to reverberation (see bottom panels of Figure 2).

While the results in Figure 2 show the effects of reverberation on ITD only, similar effects arise for ILD and spectral shape cues; i.e., the primary effects of reverberation are to distort localization cues, and this distortion increases with source azimuth and distance as well as with the number of reflecting surfaces near the listener [23]. In addition, while the results reported in Figure 2 are from a manekin, similar measurements [23, 27] show that KEMAR results fall within the range of data taken with human subjects. Overall, these results suggest that listener location may not have a large impact on the average perceived location of a sound source, but that the reliability of judgments of location may degrade with reverberation.

3. SPATIAL HEARING

Based on acoustical measurements (like those shown in Figure 2), we did not expect to see large effects of reverberation on the perceived source location for sources within a meter of the listener when the listener is in the center of the classroom. In these conditions, D/R is relatively large, and the distortion of the acoustic cues for localization relatively small and random.

Our initial study of localization in a room [3, 28, 29] replicated an earlier study of three-dimensional sound localization of nearby sources in anechoic space [30] in a classroom. Subjects pointed to the perceived sound source location while in the ***center*** of the room, far from any reflective surfaces, and then repeated the experiment with a large ***wallboard*** positioned to their side. We hypothesized that in the ***center*** condition localization performance would be nearly identical to that seen in the prior anechoic study; however, we expected performance to be worse when the wallboard was close to the listener. We found that distance localization was much better for both ***center*** and ***wallboard*** conditions than for the anechoic condition. For directional localization, response variability was slightly larger in both reverberant conditions than in anechoic conditions [3, 28, 29, 31], but there was no systematic effect on signed localization error (which was near zero). Even more surprising, we found that response variability was generally smaller in the ***wallboard*** than the ***center*** condition. Further analysis suggested that this result was due to "room learning," or an improvement in localization reliability with experience, that occurred during the ***center*** trials (which all subjects performed prior to the ***wallboard*** trials).

These findings are illustrated in Figure 3 for localization in elevation (results are similar for the left-right and distance dimensions) [29, 31]. In the figure, the mean, unsigned error (which, for these experiments, primarily reflects response variability since mean signed error is near zero) is plotted for each listener for the first and last 200 trials in the ***center*** and ***wallboard*** conditions (right side of figure) as well as for the original anechoic study (left side of figure). In the anechoic condition [30], there is no statistically-significant change in localization performance with experience; however, mean unsigned error decreases significantly with practice in the ***center*** condition. Furthermore, there is no statistically-significant change in performance when the ***wallboard*** is first put in place and no improvement between the beginning and end of the ***wallboard*** trials. These results indicate that performance improves with experience in a reverberant, but not anechoic, room (even without explicit feedback). Further, learning generalizes to conditions that differ acoustically, like conditions (e.g., from ***center*** to ***wallboard*** conditions).

Because this study confounds effects of "room learning" with acoustic effects due to the wallboard, we conducted a follow up study [27]. Using the same basic procedures, subjects performed four separate experimental sessions, each performed with the listener in a different location in the room (the locations shown in Figure 1:***left, corner, wall,*** and ***back)***.

We hypothesized that both the acoustic differences across the conditions and the room learning effect would influence results. Acoustical measurements show that distortions of spatial cues are greatest in the ***corner*** and smallest in the ***center*** locations [23, 24, 27]. Two subject groups performed the experiment; Group A performed the four sessions in the order ***center, back, left, corner***, and Group B performed the sessions in the reverse order (***corner, left, back, center***). We expected subjects from Group B, who started in the most difficult and ended in the easiest acoustic condition, to show a large improvement in localization from sessions 1 to 4. We expected subjects from Group A to show smaller changes from session 1 to session 4 because acoustic and room-learning effects opposed one another.

Figure 4 shows the variance in responses for individual subjects and the across-subject group average for both the left-right and distance dimensions. Subjects in Group B (who ended in the easiest acoustic condition) showed a large decrease in response variance from session 1 to session 4. In contrast, subjects in Group A, who heard the easiest acoustic condition first, showed no systematic change across sessions.

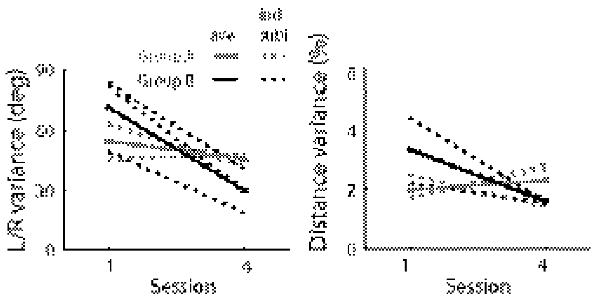

Figure 4: *Variance in localization judgments as a function of experimental session for left-right (left) and distance (right). Group A started in the easiest acoustic condition and progressed to the most difficult; Group B did the reverse.*

These results support the idea that the acoustics at different listening positions in a room influence performance, with performance generally worse for the ***corner*** location than for the ***center*** location. However, the fact that performance also improves from session to session supports the conclusions of the earlier experiment, i.e., that there is some room learning in occurs when listening in reverberant settings, an that whatever subjects learn in one room position at least partially transfers to other room positions.

These experiments show that moderate amounts of reverberation have a relatively modest impact on sound localization, causing an increase in the variability of judgments of source direction and improving the ability to judge distance. Even more, listeners appear to adapt to the reverberation present in a particular room, becoming more and more accurate at localization with experience.

4. SENSITIVITY TO REVERBERATION PATTERN

In addition to providing distance information, reverberation provides listeners with a sense of the room itself. For instance, a listener can easily tell the difference between the acoustic effects of being in a small, tiled bathroom and being in a large, heavily-carpeted living room. In fact, much of the work in the field of architectural acoustics focuses on determining what kind of room reverberation is subjectively desirable in a concert-hall setting. Whereas the usual goal in architectural acoustics is to determine what constitutes a “good” acoustic space for the average listener in the environment, we decided ask a slightly different question and measure sensitivity to differences in reverberation with changes in listener location in one particular room.

Listeners were asked to identify their room location from headphone simulations of sources at different azimuths and distances relative to the head. We hypothesized that the ability to judge room location would improve with source distance because the relative level of the reverberant energy in the signals reaching the listener increases with source distance. We expected source azimuth to have some impact because the levels and pattern of direct and reverberant energy also change with azimuth.

For each listener location, HRTFs were measured for nine source positions (all combinations of azimuths 0, 45, and 90º to the right and distances 0.15, 0.40, and 1 m). Noise samples were convolved with the set of 36 HRTFs (9 source locations

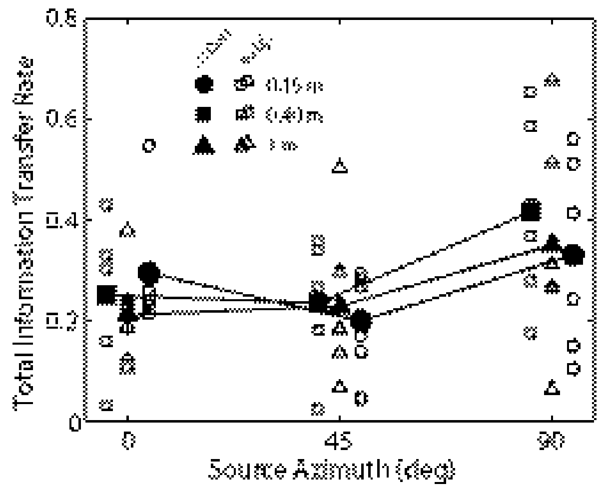

Figure 5: *Information transfer rate* ***T*** *as a function of source azimuth for three source distances (Experiment 1). Across-subject means shown by solid lines (with standard error bars); open symbols are individual subject results.*

x 4 listener locations) to generate binaural stimuli. In order to remove gross intensity differences, these stimuli were normalized so that the right-ear signal in each binaural pair (i.e., the louder of the signals in the binaural pair for the tested source locations) had the same RMS value. On each trial, the listener identified the room location, after which they were provided with correct-answer feedback.

Subjects performed 36 blocks of trials. In each block, all trials simulated the same source azimuth and distance; differences from trial to trial were due primarily to changes in the reverberation pattern. Each block consisted of 32 trials (8 presentations from each of the 4 room locations, in random order). Prior to each block of 32 trials, subjects could listen to presentations from each of the room locations as many times as they wished; testing began only when subjects felt ready to proceed. Subjects performed 12 blocks per experimental session, during which the source distance was held constant. The three source azimuths were presented in random order within each session (different orders for each day and subject), so that each day, a subject heard four consecutive blocks from the same source location followed by two sets of four blocks, each simulating a different source azimuth. To reduce any artifacts due to training, the first block in each condition are not analyzed in the results reported here.

Because we were primarily interested in how well subjects could tell listener locations apart, we analyzed *T,* the information transfer ratio, for each source location and subject (see [32]). In theory, *T* ranges between zero and one and is exactly equal to one if knowing the subject’s response perfectly predicts the stimulus presented. Low values of *T* arise when responses are independent of the stimulus. Figure 5 shows that in this experiment performance is relatively poor and individual differences large (e.g., *T* ranges from 0.1 – 0.7 for 90˚ sources). There is a modest effect of azimuth: performance is slightly (but significantly) better for sources at 90°; however, distance caused no statistically-significant main effect on *T* (multi-way ANOVA analysis, $p < 0.05$).

In order to gain further insight into what the listeners could hear, *T* was analyzed for each pair of room locations to

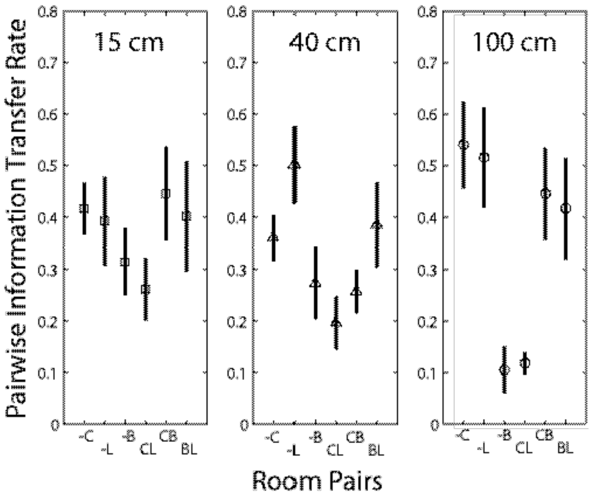

Figure 6: *Pairwise information transfer rate for all combinations of room locations averaged across subjects (with across-subject standard error). Room locations: -: **center**; C: **corner**; B: **back**; L: **left**.*

determine which pairs were relatively difficult to tell apart and relatively easy. *T* was computed for each room pair for each listener and source location. For each subject, these values were averaged across azimuth (which had little effect on the pairwise *T*).

Figure 6 shows the across-subject mean and standard error of *T* for all room pairings and for each of the source distances. These results show that for nearby sources, all room pairings were roughly equally discriminable; however, as distance increased, performance for four room pairs increased while for two pairs it decreased. More specifically, at the 1 m source distance, subjects could not distinguish between the two locations in which neither ear faced a wall (***center*** and ***back***) or between the two locations in which the left ear faced a wall (***corner*** and ***left*** positions); however, they rarely made confusions across these categories (e.g., they rarely confused ***center*** and ***corner*** locations).

A follow-up study comparing monaural and binaural performance suggests that listeners are actually better at judging the simulated listener location in the room when listening monaurally to the head-shadowed ear signal [33]. In other words, normal, binaurally listening actually decreases a listener's ability to discriminate between different patterns of reverberation. This finding hints that monaural (possibly spectral) cues are the most salient acoustic cue that could be used in this task, but that binaural listening degrades the ability to use these cues, as if the system automatically combines the signals at the two ears in a way that factors out reverberation effects.

These results suggest that under normal listening conditions listeners cannot easily discriminate where they are in an everyday room from the signals reaching their ears. For the most distant sources tested, listeners have a modest ability to discriminate between room locations in which one ear faces a wall and those in which neither ear is facing a wall, but cannot discriminate within these categories. For nearer sources, listener have some ability to discriminate across all combinations of room locations, but are not particularly good at discriminating any room pairs. It appears that the presence of very early, strong echoes (as occur when one ear faces the wall and the source is relatively distant) are easy to hear, but that other differences are difficult to discriminate. More generally, this study suggests that listeners are insensitive to many of the fine details in the pattern of echoes reaching the ears, as previously hinted by previous studies [9, 34-37]. In practical terms, the fact that listeners cannot reliably identify differences between different listener locations in a room demonstrates that simplified reverberation models may be sufficient for many spatial auditory displays.

5. SPEECH INTELLIGIBILITY AND SPATIAL UNMASKING OF SPEECH

Spatial hearing is not the only aspect of auditory perception influenced by reverberation; reverberation can have a dramatic impact on the temporal modulations in a signal. The signal reaching the ears is mathematically equivalent to the "clean" source signal convolved with the reverberant impulse response of the head and room. Because the room impulse often has a long temporal extent, there is a general tendency for reverberation to smear-out the energy in the original signal and reduce any envelope modulations present in the original signal [38, 39]. Unfortunately, the temporal modulations in speech are one of the primary sources of information about speech content; as a result, reverberation can degrade speech intelligibility [38, 40, 41].

In many cases a source of interest (the target) is heard at the same time as a competing sound (the masker). It is well known that in anechoic conditions, spatial separation of target and masker improves speech intelligibility. This effect (known as "spatial unmasking") arises both from changes in the target and masker energy levels at the ears with spatial separation and from neural processing effects. In particular, spatial separation of target and masker generally increases the target-to-masker ratio (TMR) at one ear (the "better ear") and decreases it at the other ear, leading to simple improvements in speech intelligibility due to changes in the TMR. However, listeners often perform better when listening to binaural presentations of spatially-separated target and masker than when listening to monaural presentations of the better-ear signal (e.g., see [42-48]), suggesting that binaural processing leads to further improvements.

Studies have shown that reverberation can decrease the contribution of binaural processing to spatial-unmasking [49-53]. However, in many of these studies, the reverberation presented was unnatural in its level and/or in its structure. Additionally, most of these studies were conducted using methods that make it difficult to tease apart what factors contributed to the observed results.

We have performed some preliminary studies to determine how moderate levels of reverberation (found in everyday rooms) influence both speech intelligibility and spatial unmasking of speech. We have shown [54] that the modest levels of reverberation arising at the ears of the listener in the ***center*** room position do not significantly degrade sentence intelligibility; indeed, in some cases, sentences presented with reverberation are more audible (and hence more intelligibility) than sentences presented without reverberation. In addition, these levels of reverberation do not destroy spatial unmasking effects, at least for the spatial

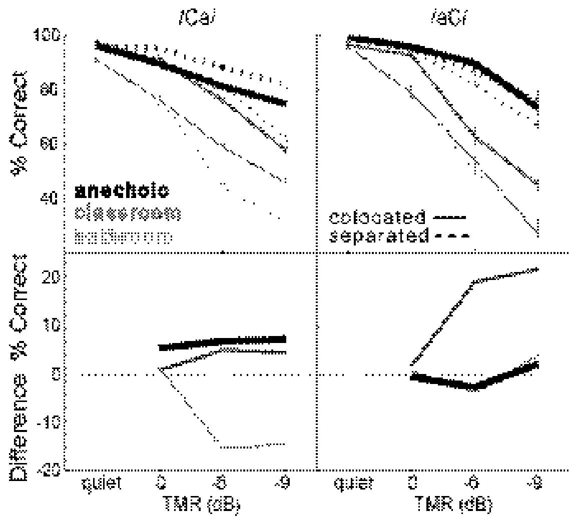

Figure 7. *Monaural performance for (a, c) initial and (b, d) final consonants. Bottom panels show effect of spatial separation (positive values indicate improvements with spatial separation).*

configurations of target and masker we tested in our initial studies [19].

More recently [20], we examined spatial unmasking of speech tokens consisting of CV and VC (V= /a/) tokens. Listeners performed a one-interval, nine-alternative, forced-choice experiment in which they identified which of the nine obstruent consonants /b,d,g,p,t,k,f,v,dh/ was presented, either in quiet or in the presence of a speech-shaped noise masker (equal to the average spectra of all target speech tokens). Five normal-hearing subjects were tested on both initial and final consonant identification.

KEMAR HRTFs were used to simulate the target and masker at different spatial locations in three different acoustic environments (***anechoic*** and ***center*** conditions as in Figure 2, as well as a ***bathroom*** condition using HRTFs from a tiled bathroom that is roughly 4 m x 2.5 m x 3.5 m). Both target utterances and noise masker were simulated at a distance of 1 m. The target was always simulated from in front of the listener (0°); the masker (when present) was simulated from either 0° or 45° to the right of the subject. Identification performance was measured as a function of TMR at the acoustically "better ear" to estimate the psychometric function. Subjects were tested binaurally and monaurally (better ear only) in quiet, with the masker in front of the listener and with the masker at 45˚.

Figure 7 plots percent-correct identification scores for the monaural test conditions. The top panel plots percent correct as a function of TMR at the better ear (note that chance performance is 1/9 or 11%). The bottom panel plots the difference between performance for spatially-separated and spatially-coincident target and masker (note change in vertical scale). Because the simulated energy emitted from M was adjusted to fix the overall TMR to the desired value at the better ear, the effects of monaural spatial unmasking are reduced compared to what would happen if the simulated masker were displaced in location (i.e., we normalized the signals to remove gross level changes due to moving the masker). This normalization was undertaken in order to emphasize the spatial unmasking effects in which we were most interested.

In general, the effect of the room reverberation was modest when listening in quiet; listeners were able to perform essentially equally well in all three room conditions. There is a consistent trend for performance to decrease with decreasing TMR. Furthermore, there is a strong interaction between reverberation and TMR; performance decreases with TMR most rapidly in the bathroom and least rapidly in anechoic space.

Much of the information about consonant identity is conveyed by acoustic cues in the 2 kHz region of the spectrum (e.g., see [48]). For a source off to the side, the head attenuates energy at frequencies above 1.5 kHz so that the TMR is generally frequency dependent. For the target and masker locations in the current study, the TMR is larger at higher frequencies than at lower frequencies when target and masker are spatially separated. Thus, even though the root-mean-square TMR is normalized, the TMR in the critical frequency region between 1.5- 5 kHz improves when target and masker are spatially separated. In the anechoic condition, improvements in TMR in this critical frequency region with spatial separation of target and masker may explain the observed improvement in performance for initial consonants in the separated condition. In the classroom there is consistent spatial unmasking due to changes in the TMR in the critical frequency region for both initial and final consonants. In the bathroom there is no spatial unmasking for final consonants; furthermore, for initial consonants there is actually 'spatial masking:' performance is *worse* when target and masker are spatially separated than when they are at the same location.

Figures 8 and 9 compare scores for binaural and monaural conditions. Although binaural performance is generally better than monaural performance, similar improvements occur when the target and masker are at the same locations and when they are in different location. In other words, there is essentially no *spatial* unmasking beyond monaural effects due to changes in TMR in the critical frequency region around 2 kHz. Although binaural processing does not contribute to the spatial unmasking of the tested consonants, binaural performance is generally better than monaural in both reverberant environments, regardless of whether target and masker are at the same or different locations. This finding is consistent with some previous studies of binaural and monaural speech discrimination of reverberant speech [50] and may be due to a statistical decorrelation of the signals at the two ears. Essentially, the two ear signals differ because the reverberation reaching the ears differs; thus, the two ear signals may effectively provide the listener with two independent looks at target and masker. The observed binaural advantage is very different from the binaural advantages normally discussed in the literature, as it does not appear to be due either to explicit comparisons between the signals at the two ears [48] or to attending to one particular spatial location [55-58].

The results of this study differ from those of our own previous studies using nearly identical procedures with *sentences* for targets, rather than consonants [19, 54, 59]. In these previous studies, we found significantly binaural contributions to speech intelligibility, but only when target and masker were spatially separated. Of course, there are a number of additional acoustic and conextual (e.g. lexical,

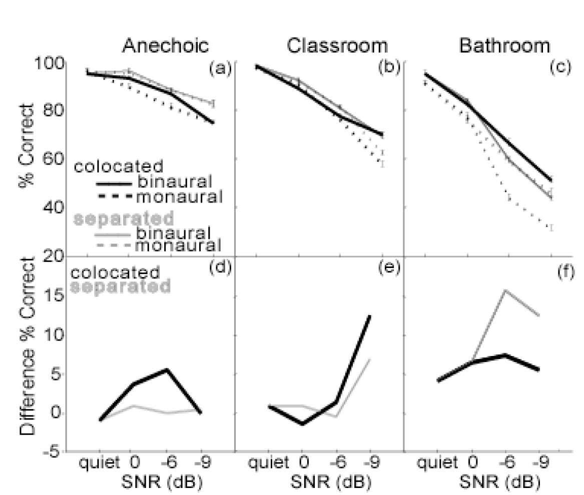

Figure 8. *Performance for initial consonants. (a, d) anechoic, (b, e) classroom, and (c, f) bathroom. Bottom panels show binaural advantage (binaural - monaural).*

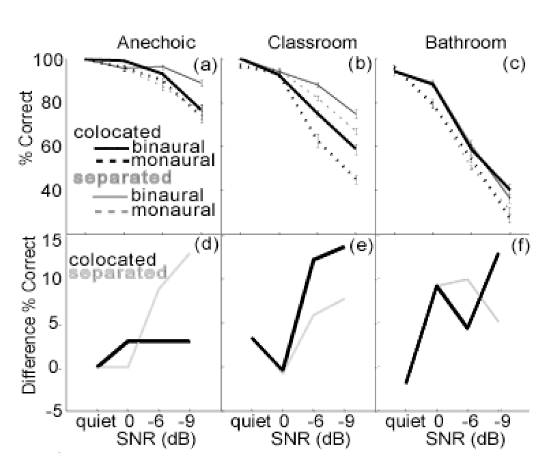

Figure 9. *Performance for final consonants. (a, d) anechoic, (b, e) classroom, and (c, f) bathroom. Bottom panels show binaural advantage (binaural – monaural).*

syntactic) cues available in a sentence perception task compared to in a consonant identification task.

Taken together, these studies suggest that binaural processing advantages are present when listening in reverberant environments, and that the nature of these binaural advantages depends on the type of stimuli presented. For sentence materials, there are spatial-separation advantages that appear to be mediated by "traditional" binaural/spatial processing. However, for phoneme recognition, binaural contributions do not come about from spatial separation of target and masker, but from some other mechanism.

Results from studies of this type can be used guide the design of spatial auditory displays by helping to determine what amount of reverberation can be included (to improve realism, provide distance cues, etc.) without perceptually degrading the source signal or destroying important spatial unmasking effects.

6. SUMMARY

Acoustical analysis shows that ordinary room reverberation distorts all aspects of the signals reaching a listener's ears. The acoustic effects of reverberation are significant even when the source is relatively close to the listener. However, the various perceptual studies reviewed here show that the human listener not only copes well with reverberation, but also even benefits from its presence in many cases.

Moderate levels of reverberation are critical for achieving a subjectively-realistic simulation of sound sources in three-dimensional space, increasing the externalization of the simulated sources and improving the ability of the listener to judge source distance. The reverberation levels arising in a moderate-sized room have only a minor degrading effect on directional hearing (most notably, by increasing response variability). In addition, experience in the room leads to improved localization performance even in the absence of feedback, as if listeners automatically and naturally adapt and adjust to reverberation and calibrate their perception to the levels of reverberation they hear. While listeners are sensitive to gross characteristics of the reverberation pattern reaching the ears, they are not particularly adept at discriminating between the exact timing and direction of the echoes reaching the ears. Finally, to a first-order approximation, the effect of moderate reverberation on speech intelligibility is to improve audibility of the speech signal without destroying the contribution of binaural processing and spatial separation. However, the combined effects of a competing source and reverberation do cause degradations that are worse than might be expected from independently considering the effects of noise and reverberation.

Further experiments of the type outlined here will help to determine how reverberation in natural environments affects both the signals reaching the ears and the processing of sound by the human observer. By using virtual spatial auditory displays to tease apart the influence of reverberation on perception, we will be able to design effective spatial auditory displays that provide realistic spatial percepts without degrading the information conveyed by the sound sources in the auditory display.

7. ACKNOWLEDGEMENTS

This project was supported by grants from the Air Force Office of Scientific Research, the National Science Foundation, and the Alfred P. Sloan Foundation. A number of students and colleagues assisted in collecting and analyzing the data reviewed herein, including Tara Brown, Doug Brungart, Scarlet Constant, Nat Durlach, Sasha Devore, Kosuke Kawakyu, Norbert Kopco, and Suraj Ram.

8. REFERENCES

[1] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," *Annual Review of Psychology*, vol. 42, pp. 135-159, 1991.

[2] F. L. Wightman and D. J. Kistler, "Sound localization," in *Human Psychophysics*, W. A. Yost, A. N. Popper, and R. R. Fay, Eds. New York: Springer Verlag, 1993, pp. 155-192.

[3] S. Santarelli, "Auditory Localization of Nearby Sources in Anechoic and Reverberant Environments," in *Cognitive and Neural Systems*. Boston, MA: Boston University, 2000.

[4] B. G. Shinn-Cunningham, "Localizing sound in rooms," Proceedings of the ACM/SIGGRAPH and Eurographics Campfire: Acoustic Rendering for Virtual Environments, Snowbird, Utah, 2001.

[5] D. H. Mershon and L. E. King, "Intensity and reverberation as factors in auditory perception of egocentric distance," *Perception and Psychophysics*, vol. 18, pp. 409-415, 1975.

[6] D. H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan, "Effects of room reflectance and background noise on perceived auditory distance," *Perception*, vol. 18, pp. 403-416, 1989.

[7] N. I. Durlach, A. Rigapulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel, "On the externalization of auditory images," *Presence*, vol. 1, pp. 251-257, 1992.

[8] D. R. Begault, "Auditory and non-auditory factors that potentially influence virtual acoustic imagery," Proceedings of the Audio Engineering Society International Conference on Spatial Sound Reproduction, Rovaniemi, Finland, 1999.

[9] D. R. Begault, E. M. Wenzel, A. S. Lee, and M. R. Anderson, "Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, pp. 904-916, 2001.

[10] D. R. Begault, "Binaural auralization and perceptual veridicality," *Journal of the Audio Engineering Society*, vol. preprint 3421, 1992.

[11] D. R. Begault, "Perceptual effects of synthetic reverberation on three-dimensional audio systems," *Journal of the Audio Engineering Society*, vol. 40, pp. 895-904, 1992.

[12] W. L. Martens and A. Yoshida, "Psychoacoustically-based control of auditory range: Display of virtual sound sources in the listener's personal space," Proceedings of the International Conference on Information Society in the 21st Century: Emerging Technologies and New Challenges (ISO2000), Aizu-Wakamatsu, Japan, 2000.

[13] W. L. Martens, "Pseudophonic listening in reverberant environments: Implications for optimizing auditory display for the human user of a telerobotic system," presented at International Conference on Information Society in the 21st Century: Emerging Technologies and New Challenges, Aizu-Wakamatsu, Japan, 2000.

[14] W. L. Martens, "Psychophysical calibration for controlling the range of a virtual sound source: Multdimensional complexity in spatial auditory display," Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, 2001.

[15] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *Journal of the Acoustical Society of America*, vol. 11, pp. 1832-1846, 2002.

[16] A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, pp. 517-520, 1999.

[17] B. E. D. Kingsbury, N. Morgan, and S. Greenberg, "Improving ASR performance for reverberant speech," Proceedings of the ESCA workshop on Robust Speech Recognition, Pont-a-Mousson, 1997.

[18] K. S. Helfer, "Binaural cues and consonant perception in reverberation and noise," *Journal of Speech and Hearing Research*, vol. 37, pp. 429-438, 1994.

[19] B. G. Shinn-Cunningham, "Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays," Proceedings of the International Conference on Auditory Displays, Atlanta, GA, 2002.

[20] S. Devore and B. G. Shinn-Cunningham, "Perceptual consequences of including reverberation in spatial auditory displays," Proceedings of the International Conference on Auditory Displays, 2003.

[21] B. G. Shinn-Cunningham, "Creating three dimensions in virtual auditory displays," in *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality*, M. Smith, G. Salvendy, D. Harris, and R. Koubek, Eds. New Jersey: Lawrence Erlbaum, 2001, pp. 604-608.

[22] B. G. Shinn-Cunningham, "Distance cues for virtual auditory space," Proceedings of the IEEE-PCM 2000, Sydney, Australia, 2000.

[23] T. J. Brown, "Characterization of Acoustic Head-Related Transfer Functions for Nearby Sources," in *Electrical Engineering and Computer Science*. Cambridge, MA: Massachusetts Institute of Technology, 2000.

[24] B. G. Shinn-Cunningham, J. G. Desloge, and N. Kopco, "Empirical and modeled acoustic transfer functions in a simple room: Effects of distance and direction," Proceedings IEEE-WASPAA, New Pfalz, New York, 2001.

[25] B. G. Shinn-Cunningham, "Creating three dimensions in virtual auditory displays," presented at Proceedings of HCI International 2001, New Orleans, 2001.

[26] N. Kopco and B. G. Shinn-Cunningham, "Auditory localization in rooms: Acoustic analysis and behavior," Proceedings of the 32nd International Acoustics Conference -- EAA Symposium, Zvolen, Slovakia, 2002.

[27] N. Kopco and B. G. Shinn-Cunningham, "Effect of listener location on localization cues and localization performance in a reverberant room," Mid-Winter meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL, 2001.

[28] S. Santarelli, N. Kopco, B. G. Shinn-Cunningham, and D. S. Brungart, "Near-field localization in echoic rooms," *Journal of the Acoustical Society of America*, vol. 105, pp. 1024, 1999.

[29] S. Santarelli, N. Kopco, and B. G. Shinn-Cunningham, "Localization of near-field sources in a reverberant room," Mid-Winter meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL, 1999.

[30] D. S. Brungart and N. I. Durlach, "Auditory localization of nearby sources II: Localization of a broadband source in the near field," *Journal of the Acoustical Society of America*, vol. 106, pp. 1956-1968, 1999.

[31] B. G. Shinn-Cunningham, N. Kopco, and S. Santarelli, "Computation of acoustic source position in near-field listening," presented at 3rd International Conference on Cognitive and Neural Systems, Boston, MA, 1999.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[33] B. G. Shinn-Cunningham and S. Ram, "Identifying where you are in a room: Sensitivity to room acoustics,"

Proceedings of the International Conference on Auditory Display, 2003.
[34] S. Bech, "Audibility of individual reflections in a complete sound field II," *Journal of the Acoustical Society of America*, vol. 97, pp. 3320, 1995.
[35] S. Bech, "Timbral aspects of reproduced sound in small rooms. I," *Journal of the Acoustical Society of America*, vol. 97, pp. 1717-1726, 1998.
[36] S. Bech, "Timbral aspects of reproduced sound in small rooms. II," *Journal of the Acoustical Society of America*, vol. 99, pp. 3539-3549, 1998.
[37] D. R. Begault, B. U. McClain, and M. R. Anderson, "Early reflection thresholds for virtual sound sources," Proceedings of the 2001 International Workshop on Spatial Media, Aizu-Wakamatsu, Japan, 2001.
[38] T. Houtgast, H. J. M. Steeneken, and R. Plomp, "Predicting speech intelligibility in rooms from the modulation transfer function I. General room acoustics," *Acustica*, vol. 46, pp. 60-72, 1980.
[39] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *Journal of the Acoustical Society of America*, vol. 77, pp. 1069-1077, 1985.
[40] M. ter Keurs, J. M. Festen, and R. Plomp, "Effect of spectral envelope smearing on speech reception: I," *Journal of the Acoustical Society of America*, vol. 91, pp. 2872-2880, 1992.
[41] R. Plomp, H. J. M. Steeneken, and T. Houtgast, "Predicting speech intelligibility in rooms from the modulation transfer function II. Mirror image computer model applied to rectangular rooms," *Acustica*, vol. 46, pp. 73-81, 1980.
[42] A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *Journal of the Acoustical Society of America*, vol. 83, pp. 1508-1516, 1988.
[43] A. W. Bronkhorst and R. Plomp, "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," *Journal of the Acoustical Society of America*, vol. 92, pp. 3132-3139, 1992.
[44] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acustica*, vol. 86, pp. 117-128, 2000.
[45] R. Drullman and A. W. Bronkhorst, "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," *Journal of the Acoustical Society of America*, vol. 107, pp. 2224-2235, 2000.
[46] M. L. Hawley, R. Y. Litovsky, and H. S. Colburn, "Speech intelligibility and localization in a multi-source environment," *Journal of the Acoustical Society of America*, vol. 105, pp. 3436-3448, 1999.
[47] J. Peissig and B. Kollmeier, "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *Journal of the Acoustical Society of America*, vol. 101, pp. 1660-1670, 1997.
[48] P. M. Zurek, "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, G. Studebaker and I. Hochberg, Eds. Boston, MA: College-Hill Press, 1993.
[49] S. A. Gelfand and S. Silman, "Effects of small room reverberation on the recognition of some consonant features," *Journal of the Acoustical Society of America*, vol. 66, pp. 22-29, 1979.
[50] S. A. Gelfand and I. Hochberg, "Binaural and monaural speech discrimination under reverberation," *Audiology*, vol. 15, 1976.
[51] J. F. Culling, Q. Summerfield, and D. H. Marshall, "Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels," *Speech Communication*, vol. 14, pp. 71-95, 1994.
[52] C. J. Darwin and R. W. Hukin, "Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention," *Journal of the Acoustical Society of America*, vol. 108, pp. 335-342, 2000.
[53] C. J. Darwin and R. W. Hukin, "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," *Journal of the Acoustical Society of America*, vol. 107, pp. 970-977, 2000.
[54] B. G. Shinn-Cunningham, S. Constant, and N. Kopco, "Spatial unmasking of speech in simulated anechoic and reverberant rooms," Mid-Winter meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL, 2002.
[55] R. L. Freyman, U. Balakrishnan, and K. Helfer, "Release from informational masking in speech recognition," Mid-Winter Meeting of the Association for Research in Otolaryngology, St. Petersberg Beach, FL, 2000.
[56] R. L. Freyman, K. S. Helfer, D. D. McCall, and R. K. Clifton, "The role of perceived spatial separation in the unmasking of speech," *Journal of the Acoustical Society of America*, vol. 106, pp. 3578-3588, 1999.
[57] D. S. Brungart, "Informational and energetic masking effects in the perception of two simultaneous talkers," *Journal of the Acoustical Society of America*, vol. 109, pp. 1101-1109, 2001.
[58] D. S. Brungart, "Evaluation of speech intelligibility with the coordinate response method," *Journal of the Acoustical Society of America*, vol. in press, 2001.
[59] B. G. Shinn-Cunningham, L. Mraz, and N. Kopco, "Effect of reverberation on spatial unmasking for nearby speech sources," *Journal of the Acoustical Society of America*, vol. 109, pp. 2468, 2001.