CHAPTER 1

NORMAL LISTENING IN TYPICAL ROOMS

THE PHYSICAL AND PSYCHOPHYSICAL CORRELATES OF REVERBERATION

David A. Berkley
Jont B. Allen

This chapter makes four major points:

1. Room reverberation may be characterized by two distinct perceptual components that have been termed *coloration* and *echo*. Under quiet conditions, listening preference for reverberant speech is dependent on both of these components.
2. The corresponding physical variables are the reverberation time T_{60} and the talker-listener distance. Both of these physical variables play an important role in listener preference for a given listening condition.
3. Computer simulations have been used to study coloration and echo perception for normal hearing subjects under quiet (unmasked) conditions.
4. The optimal-preference listening condition depends on a simple trade-off between the coloration and the echo components. A method of measuring this listener preference will be described. A mapping between the physical variables and the perceptual variables allows us the prediction of preference for a given listening condition. This preference is different for one- and two-ear listening; the difference may be defined as the *dichotic release from reverberation.*

PHYSICAL BASIS OF REVERBERATION

Definitions

The term *normal listening,* as used in this chapter, is defined as the listening condition for a person with normal hearing thresholds in both ears listening to full bandwidth speech with either one or two ears in a reasonably quiet (unmasked) condition. The term *typical rooms* is used here to mean rooms and spaces of the home and daily working environment.

Intelligibility Versus Preference

Under these normal listening conditions, it has been found that reverberation does not significantly reduce the intelligibility of the perceived speech. Of course, there are conditions for which reverberation does reduce speech intelligibility, and these are discussed by Nábělek in Chapter 2. However, when intelligibility is not reduced, it follows that an alternative perceptual measure is necessary to quantify these reverberant effects. In this chapter, both *differences* between reverberant conditions and *preference* are used as this perceptual measure.

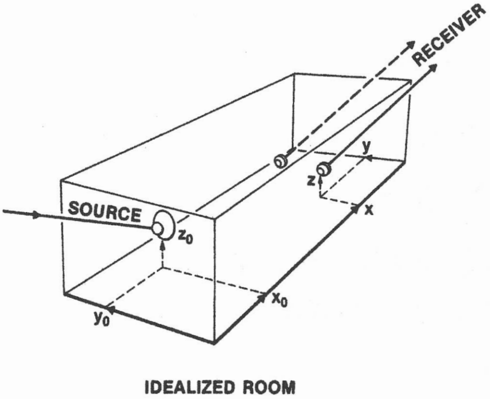

FIGURE 1.1. Idealized rectangular solid "room."

Ideal Rooms

For the purpose of this study, the room is treated as a rectangular, solid enclosure, as shown in Figure 1.1. In this idealization, the effects of absorption and diffusion of sound produced by any objects in the room are ignored. For the model room, all absorption must take place at the walls. Experiments have shown that this idealized model produces both realistic-sounding reverberation and other physical phenomena (e.g., modal densities, reverberation decays) that agree with measurements on real rooms (Allen and Berkley, 1979). Speech sounds are introduced into the model room at a "point" sound source, which might be a person's mouth, idealized to radiate sound equally in all directions. The resulting energy in the room is picked up or "heard" by a point receiver, which might be a similarly idealized "ear."

Transfer Function

How can the transmission of sound energy between source and receiver be described? Two alternative formulations of this problem have been made:

The frequency domain approach (Morse and Ingard, 1968), using a normal mode expansion

The time domain approach (Mintzer, 1950), in which the time course of reflections made by the sound waves reflecting from the room walls is followed and in which a pulse of sound at the source thus turns into many pulses at the receiver

The normal mode method has considerable theoretical attraction and is the method that has generally dominated theoretical considerations of room reverberation. However, the approach in this chapter uses the time domain (or "impulse") method which allows a much more computationally efficient approach when modeling the room transfer function. The frequency response of rooms is referred to only when it helps us to understand perception.

Single-Wall Echoes. What happens when a pulse of sound is emitted by the source? Consider first a single wall, as shown in Figure 1.2. If an acoustic pressure pulse is introduced into the space, the resulting acoustic wave will propagate away from the source, with its sound pressure level (P) attenuated as the reciprocal of distance traveled. The direct wave will intersect the microphone with level $P_1 = P/r_1$. The pulse will also reflect off the wall, as shown, being attenuated by the pressure reflection coefficient k, and finally arrive at the pickup, after traveling the longer distance r_2, with pressure $P_2 = kP/r_2$.

If opposing walls are present, this process goes on forever. The wave, bouncing around the room, is attenuated as it travels and is absorbed on each wall reflection.

Impulse Response. Figure 1.3 shows what happens in the room model of a small office about 10' × 12' × 10'. The first pulse to reach the receiving point is always the direct sound

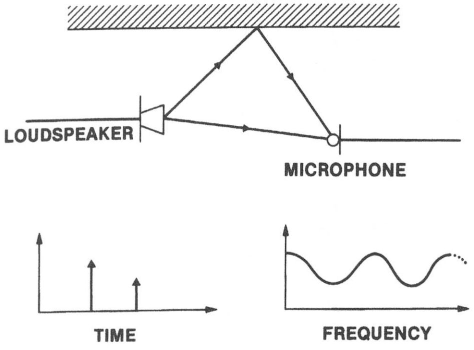

FIGURE 1.2. Sound reflected from a single wall has two paths to the receiving point. The impulse response consists of two pulses, and the frequency response varies over frequency.

from the source. The first few reflections are from nearby walls and are fairly well defined. Later reflections, representing multiple reflections from the walls, are so numerous that they blur together.

This complicated result has many uses. The pulse response contains all the information available about the room derivable from this combination of source-receiver locations. In particular, using computer methods, speech can be processed using this response so that it is the same as speech actually passed through the room. The signal processing method by which this is done is called *convolution* (Oppenheim and Schafer, 1975). Returning to the one-wall case of Figure 1.2, if the speech signal s(t) is convolved with the wall impulse response, the received signal is

$$P(t) = s(t - T_1)/r_1 + k \cdot s(t - T_2)/r_2,$$

where T_1 is the time for the direct sound to travel distance r_1 to the pickup, T_2 is the propagation time for the reflected wave traveling distance r_2, and k is the wall reflection coefficient. Therefore, to get a resulting signal output, sum the input signal with itself, where the time delay and gain of each term are given by the corresponding values of the impulse response samples. Further details of the room model implementation may be found in the appendix to this chapter.

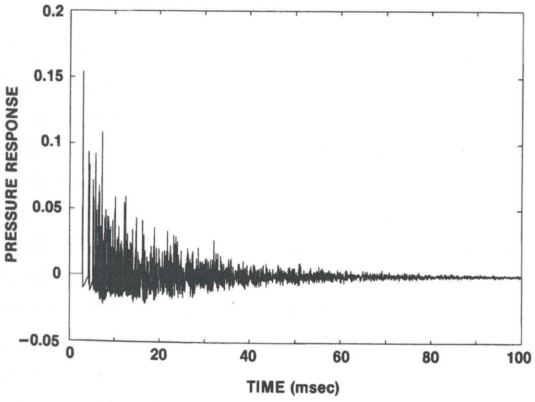

FIGURE 1.3. Impulse response of a simulated rectangular room.

PERCEPTION OF REVERBERATION

Perception for One Reflection

Echo. The physical origin and description of reverberant sound is fairly simple. How then does one *perceive* the resulting reverberant signal? Return to the simple case of a single reflection. Suppose the wall is distant from the source, perhaps 25 feet, producing approximately a 50-msec delay in the reflected pulse (sound travels about 1 foot/msec), as seen in Figure 1.4 (upper). In this case, a clear single echo is heard. When multiple reflections of this type occur, as in a large hard-walled room, we hear the resulting "reverberation" as a clutter of echoes.

Coloration. With a signal having short echo delay (i.e., when the wall is two feet away, giving approximately a 2-msec delay), the perception is one of a change in timbre of the speech, usually called coloration. When the pulses are close together, the ear cannot dis-

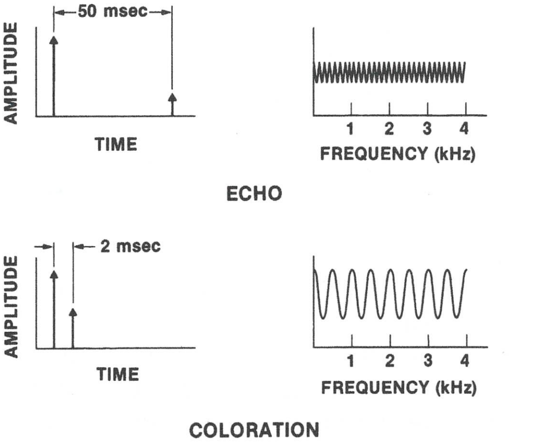

FIGURE 1.4. The impulse response (left) and frequency response (right) for a single long-time (upper) and short-time (lower) reflections.

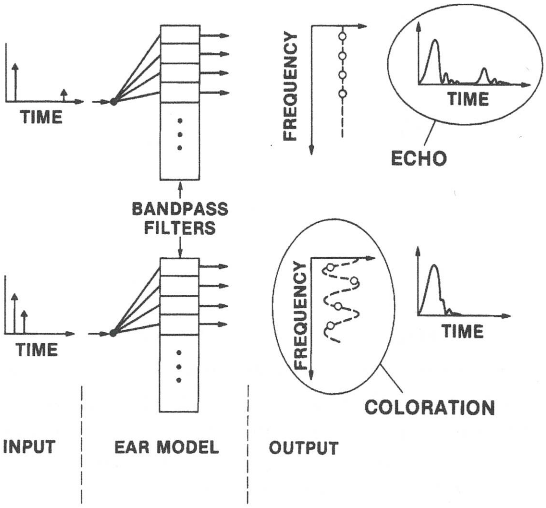

FIGURE 1.5. The effect of a long-time echo (top) and a short-time echo (bottom) on a simple hearing model.

tinguish the time difference between the pulses, and the frequency response of the pulse pair shown in Figure 1.4 (lower) is heard. Thus, for short delays, the room acts like a frequency shaping filter, distorting (or coloring) the frequency content of the original speech signal.

Ear Model. Figure 1.5, which shows a highly simplified model of the ear, illustrates why perception breaks into echo and coloration. The incoming signal is broken into frequency bands, corresponding to the ear's critical bands. These critical bands are the result of the cochlear filters, which have a memory, or duration, that is roughly inversely proportional to their bandwidth and is about 5 to 20 msec, depending on the critical band in question.

The two sample signals show why the two different types of perception are expected. The long-delay pulse (Figure 1.5, upper) produces no variation in amplitude across frequency because the filters of the ear have shorter memory than the distance between the pulses, and the time or echo nature of the signal is well preserved. In the short-delay case (Figure 1.5, lower), frequency content of the pulse pair is well-preserved in the output, but all time information is lost because the memory of the cochlear filters is greater than the distance between pulses, causing the pulses to interact within the cochlea.

Perception in Real Rooms

Intuitive Dissection. Figure 1.6 shows a simplified real-room impulse response similar to the one examined previously and a separation or dissection of the impulse into the direct (a), early (b), and late (c) portions of the response. An actual room impulse response was separated in this manner and speech was convolved with each of the three parts. This allowed listening separately to the effects of the early and late echo. The results were simple extensions of those for a single echo, as previously discussed. Windowing at 50 msec primarily produces a change in timbre or coloration of the speech, whereas lis-

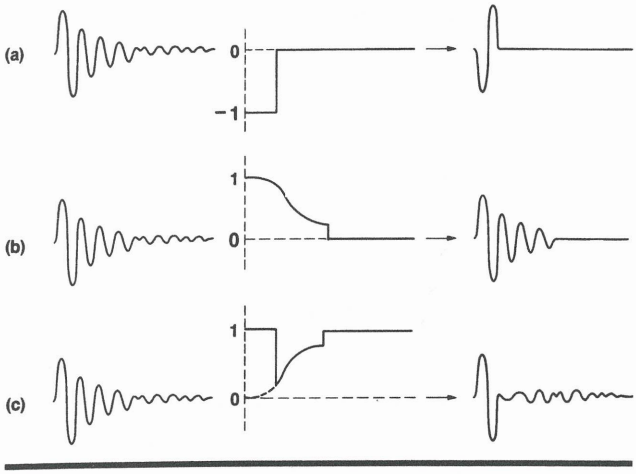

FIGURE 1.6. The effects of the short and long times may be heard if any impulse response is dissected into (a) the direct sound, (b) early reflections (<50 msec), and (c) late reflections (>50 msec).

tening to direct sound plus the late response of the room yielded "booming" or echoing speech (Berkley, Curtis, and Allen, 1973).

Quantification of Perception. This qualitative understanding of perception of coloration and echo has been accepted for some time. The stumbling block has been how to formalize these results under "realistic" conditions. Two areas of difficulty have existed:

1. The perceptual phenomena are multidimensional (i.e., no simple psychophysical measurement method can analyze the complex underlying perceptual basis by which individuals discriminate between differing reverberant conditions).
2. The underlying physical variables are also complex and, more important, many of them are difficult, if not impossible, to control fully or even measure accurately. In real rooms, reverberant effects are often corrupted by noise or imperfect recording instruments.

The Allen-McDermott Experiment

A first step toward the resolution of these two problems was taken by one of the authors (Allen) and one of his coworkers at AT&T Bell Laboratories, Barbara McDermott. The material in the next section of this chapter is drawn from an unpublished internal report on their experiments entitled "The Perceptual Variables of Small Room Reverberation" (Allen, McDermott, and Berkley, 1979) and from some later unpublished experiments performed by the other author (Berkley) and Sheryll Berggren, which extended the Allen and McDermott results to dichotic stimuli.

Experimental Design. These experiments dealt with the first of the two major problems described above by designing the experiment and analyzing the results within the framework of procedures collectively known as multidimensional scaling (Kruskal and Wish, 1978). The second problem was solved by computer simulation of the reverberant conditions (Allen and Berkley, 1979).

Physical Variables. In these experiments, the room size was a constant: 12.5' × 15' × 16.25'. By changing the surface absorptions, the room reverberation time was varied in 5 steps from 75 msec to 480 msec. (The *reverberation time* is the time that it takes for sound in the room to decay by 60 dB after being turned off.) The talker-microphone (source-listener) distance also was varied in 5 steps from 0.63 to 10.0 feet. From the 25 possible resulting room conditions, 16 were selected, and the first 512 msec of the room impulse response was calculated using the room model. (This response was sampled at 125-μsec intervals, which corresponds to a sampling rate of 8 kHz, allowing a 4-kHz bandwidth for the simulation.) Ten different sentences, each spoken by four (two male and two female) talkers, were convolved with the 16 calculated impulse responses, producing 640 digitally reverberated samples. These were then converted to analog tape recordings having a bandwidth of 100 to 4000 Hz.

The listening test experiment consisted of three parts.

Difference Judgments. Sample tapes of all possible pairs of room conditions were played to 25 untrained normal listeners (balanced over talkers, order, and sentences). The listeners were asked to rate *how different* the two rooms were on a scale of 0 (for no difference) to 9 (for maximum difference).

Preference Ratings. In a separate experiment, the same 25 subjects listened to samples of each room condition (with different talkers and sentences) and were asked to rate the rooms on a 9-point scale with descriptive adjectives (unsatisfactory, poor, fair, good, and excellent) labeling alternate scale points.

Coloration and Echo Ratings. Finally, two "experienced" listeners (researchers who had listened critically to such signals for several years) rated all the room conditions separately according to their expert judgment of the amount of echo and coloration present.

Experimental Results

Multidimensional Analysis of Difference Judgments. An analysis of the difference judgments is shown in the *difference space* of Figure 1.7. In order to discuss the results, it is necessary to describe the multidimensional scaling (MDS) procedures.

Each point on the plot in Figure 1.7 is one of the experimental room conditions and is labeled with two numbers that correspond to the talker-listener distance, in feet, and the reverberation time, in msec. The positions in the space have been found by MDS methods so that the *distances* between pairs of points

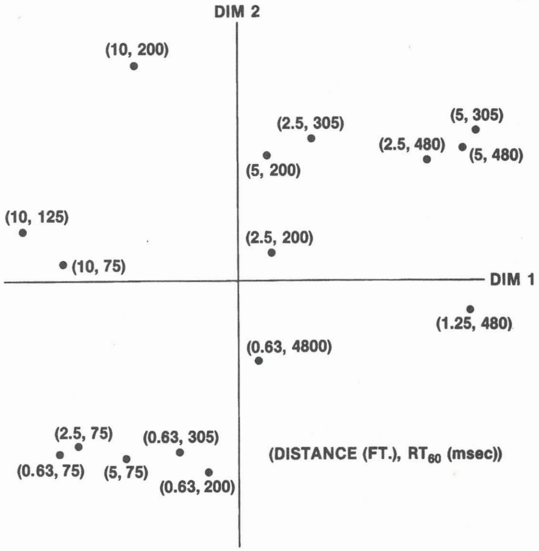

FIGURE 1.7. Two-dimensional representation of subject judgments of distance between simulated reverberation samples.

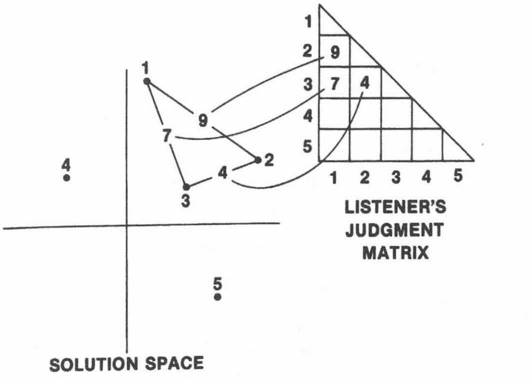

FIGURE 1.8. Example of projection of three points onto a two-dimensional solution space.

best represent the conglomerate difference judgments of the subjects. That this set of difference judgments is well represented by the two-dimensional plot is already a nontrivial result. It shows that two perceptual components make up the subjective difference judgments.

To use a classical analogy to the subjective difference judgments, consider the airline distances between cities in the United States. The multidimensional scaling programs would produce an actual two-dimensional map of the U.S. from this matrix of distances. However, if only East Coast cities were used, a one-dimensional map would be a fairly good representation. A second example is shown in Figure 1.8, where we show the relationship between three points and the listener's difference judgments in a two-dimensional space.

Projection of the Ratings. This is the beginning, rather than the end, of the required analysis. Thus the structural relationship or geometry of the underlying perceptual variables has been defined, but not the orientation or physical significance of the multidimensional space coordinates. However, given ratings of the room conditions, such as preference judgments, multidimensional linear regression can be used to *project* the measures into the multidimensional difference space. Many other measures, such as expert ratings and physical measurements on the room (e.g., the reverberation time), could also be projected in the same manner. This is done to identify the physical meaning of the difference space, to find physical measures that describe the difference space, and to help determine how the ratings functionally depend on the underlying difference space. Thus, the concept of projection is an important tool in MDS.

As an example of projection, assume the ratings to be projected are the preference data for each room condition. It is found that the room conditions define a two-dimensional perception space, based on analysis of the difference judgments. Visualize the rating for each stimulus point as a function of the two coordinates of each point in the difference space. Linear regression is then used to find a plane in the three-dimensional space that best fits the ratings as a function of the two coordinates. The projected vector is chosen in the direction of the maximum slope of the regression plane. This direction is the opposite to the steepest descent direction. In general, it is computed as a unit vector in the direction of the gradient and a length proportional to the linear regression coefficient.

Identification of the Difference Space. To identify the difference space dimensions, *assume* that the two dimensions are coloration and echo and then test this hypothesis. That this is not a bad assumption is shown by projecting the "expert" judgments of coloration and echo magnitude for each room condition into the difference space, as seen in Figure 1.9. In fact, these two perceptual variables appear to be reasonably independent since there is almost a 90-degree angle between them.

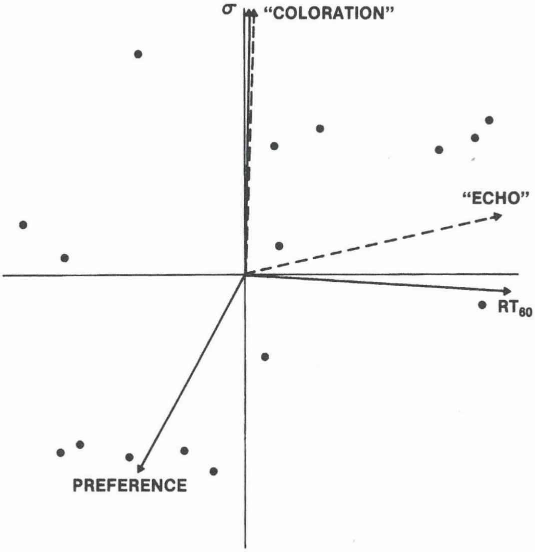

FIGURE 1.9. The difference space with superimposed vectors representing expert judgments of coloration and echo, the physical variables, T_{60}, the reverberation time, and σ, the spectral deviation (which were computed from the impulse responses), and the preference judgments. The method for projecting the vectors into the difference space is described in the text.

Relation of the Difference Space to the Physical Variables. How can these results be related to the underlying physical variables? After considering a number of possible candidates, it was found that two simple physical measures could be computed from the room impulse responses. These measures span the perception space. In other words, when these measures are projected into the difference space, they form a non-collinear set of vectors that are highly correlated to the space. The best two found were the *reverberation time T_{60}*, and the *spectral deviation σ*. The spectral deviation is a measure of the roughness of the frequency response of the room and is defined as the square root of the sum of the squares of the difference between actual room frequency response, expressed in decibels, and the average frequency response. The reverberation time and spectral deviation were found to define nearly orthogonal axes when projected into the difference space and, as seen in Figure 1.9, are well correlated with the perceptual variables, echo and coloration.

Preference. An analysis of the preference results gave an interesting surprise: Unlike the discrimination results, preference turned out to be a one-dimensional measure (established in a second MDS experiment not described here), which essentially meant that all subjects agreed on their preference for reverberant listening conditions. When projected onto the difference space, preference was found to lie between the two axes (Figure 1.9), which means that it is a function of both perceptual dimensions, coloration and echo. Hence, preference may be "predicted" from the reverberation time and spectral deviation, as shown in the linear regression equation of Figure 1.10.

Projected this way, with the preference expressed in the form of the excellent-poor scale [as a mean opinion score (MOS)], a parsimonious picture emerges of what normal diotic listeners prefer for room listening. As in Figure 1.7, the room conditions are labeled with the speaker-listener distance, in feet, and reverberation time, in msecs.

Figure 1.10 shows the results of a linear regression analysis. Allen (1982) reported a more accurate nonlinear representation for the preference $P(\sigma,T_{60})$ of the form:

$$P/P_{MAX} = 1 - 0.3\sigma T_{60},$$

where P is the preference, which may range between zero and P_{MAX}, σ is the spectral deviation, in dB units, and T_{60} is the reverberation time, in seconds. This relation assures that if either σ or T_{60} goes to zero, then the

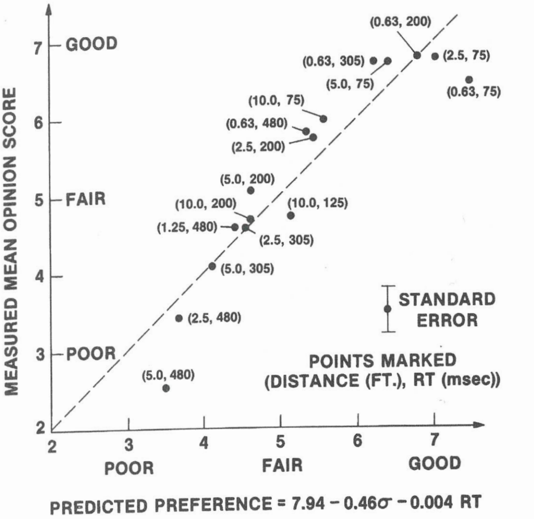

FIGURE 1.10. Preference (mean opinion score) versus predicted preference.

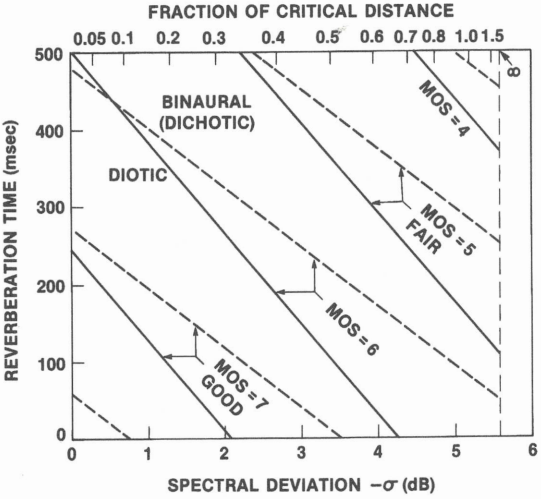

FIGURE 1.11. Iso-preference results for simulated rooms for both diotic and dichotic conditions. Note the release from reverberation that is obtained in the dichotic case relative to the diotic case.

preference will be at its maximum, which is in agreement with observation. The linear expression for the preference in Figure 1.10 can be viewed as a linearized version of the above equation. As such, it fails to account for the nonlinear interactions, which give a more accurate analysis of the preference data. (Note that the above formula obviously fails for very large reverberation times, where it predicts that P becomes negative.)

This same combination of reverberation time and spectral deviation has been replotted using the linear regression method in the form of the iso-preference curves (Figure 1.11, dashed lines). The concept of *critical distance* allows further interpretation of this figure. Critical distance is defined as that distance from source to receiver where the direct sound energy is equal to the total reverberant sound energy. The physical source-receiver distance, normalized by the critical distance, is theoretically related to spectral deviation (Jetzt, 1979), as shown on the upper scale in Figure 1.11.

Consider the example of a test room with a reverberation time of 125 msec where the critical distance is 10 feet. (This is equivalent to an office with less-than-average reverberation.) This implies that, from consideration of Figure 1.11 alone, a one-ear listener with a 4-kHz hearing bandwidth will want to be less than 2 feet away from a talker for "good" listening. The effect of natural, one-ear directionality or the influence of a directional pickup (or source) will increase the "good" distance in proportion to the increased directionality.

Binaural Release from Reverberation. In addition to the effects of directionality mentioned above, experience has also shown that two-ear listening allows greater distances for the same subjective quality. In Figure 1.11, the dashed lines show the iso-preference contours found by linear regression for a dichotic version of experiments similar to those described above (Berggren, Berkley, and McDermott, 1980). In this case, two impulse responses were computed for each room condition with the receiving points 6 inches apart. Presentations were given to the subjects dichotically, and all of the analysis was repeated as before. In the previous example, the same listener can now obtain "good" listening at about 3 feet away from the source using two ears versus 2 feet with a single ear.

SUMMARY

The physical bases for our perception of reverberation have been defined. We have found that reverberation perception is a "two-dimensional" phenomenon consisting of "coloration" and "echo" dimensions. These precepts are correlated to the physical measures, spectral deviation and reverberation time. Our preference for a given reverberant condition may be predicted from a simple function of spectral deviation (or normalized distance) and reverberation time.

APPENDIX TO CHAPTER 1

ROOM SIMULATION

The room simulation techniques used in the perceptual experiments rest on the method of images (Allen and Berkley, 1979).

Basic Physical Principles

Figure 1.12a shows the single reflection of a sound wave from a wall, shown originally in Figure 1.2, in a different form. If the wall is perfectly reflecting, it may be replaced by a second "image" source placed symmetrically with respect to the original source in analogy with an optical mirror and image. If the original source emits a sound pulse, the image emits an identical pulse at the same time. If the wall absorbs some portion of the sound wave, this may be approximately accounted for by decreasing the image strength by the wall reflection coefficient.

With two opposite, non-absorbing walls (Figure 1.12b), there are an infinite number of images, as with two opposing optical mirrors. A two-dimensional enclosure, shown in Figure 1.12c, looks still more complex, but the physics is the same. A three-dimensional structure again produces the same result, but it is not easily visually depicted. However, mathematical expressions for the receiver output with a pulse source input can be written directly for all these cases. The resulting expressions look complicated, but are only direct extensions of the equation given previously for a reflection from a single wall (Allen and Berkley, 1979).

It is remarkable that this result is exact (in the sense that it is an exact solution to the full wave equation formulation of acoustics) when the walls of the room are "hard" or reflecting perfectly. Thus, an alternative way of thinking about a room impulse response is that each sub-impulse represents the arrival of a pulse emitted by one of the images. The later is the response, the farther away is the contributing image. Even when the walls are not perfect reflectors, images still result in a good approximation of the physical results in the room.

Computer Implementation

The image representation is powerful because after a relatively short time (about one-half of the room reverberation time), the remaining incoming pulses no longer contribute sufficient energy to be perceptible or to affect other room measurements, the energy hav-

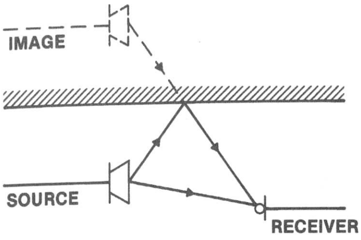

FIGURE 1.12a. The case of one wall reflection may be treated as an image source.

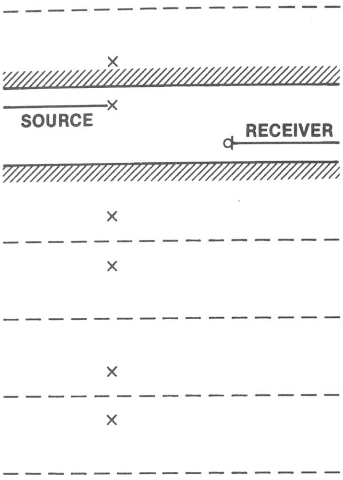

FIGURE 1.12b. Two walls becomes an infinite number of sources strung out in a line.

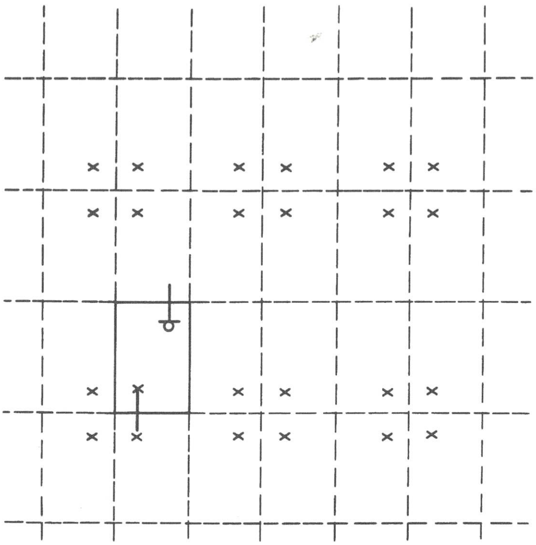

FIGURE 1.12c. In two dimensions there is a lattice of sources. A rectangular room (three dimensions) gives a three-dimensional lattice of sources much like the two-dimensional case.

ing been lost in reflections from the walls. Thus, the *significant* response is built up by summing those images within a sphere, the radius of which is determined by the distance sound travels in the significant time. Although this may still be tens of thousands of images, the computation is well suited to efficient evaluation on a modern digital computer, and a reasonably high-speed machine can compute more than 10,000 images/sec.

The other computational aspect of this problem is the convolutions, which for the cases reported on here required 8 megaflops (8 million floating-point instructions per second). For comparison, the AT&T DSP-32C digital signal processing chip is rated at 20 megaflops. This chip is now available on a PC plug-in board with digital-to-analog converters. This means that this study, if done today, could be done in real time on a PC.

The Spectral Deviation

The spectral deviation room measurement method, first worked out by John Jetzt at Bell Laboratories, was initially tested using the room simulation technique (Jetzt, 1979). Jetzt first derived the theoretical relation between spectral deviation and source-receiver distance normalized by critical distance and then verified the relationship using the room simulation. His results are shown in Figure 1.13. This verification would be impossible in a real room where the direct and reverberant energies, and thus the critical distance, cannot be accurately measured. Once verified, the spectral deviation method provides a sensitive measure for determining the critical distance, even in rooms where the reverberation time cannot be determined accurately.

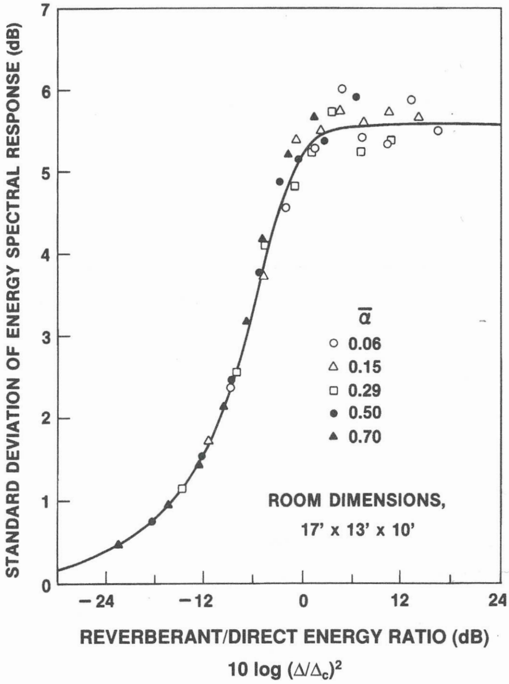

FIGURE 1.13. Jetzt (1979) first derived this plot which shows the spectral deviation σ as a function of the normalized source-receiver distance.

REFERENCES

Allen, J.B. (1982). Effects of small room reverberation on subjective preference. *J. Acoust. Soc. Am. 71,* S5.

Allen, J.B., and Berkley, D.A. (1979). Image method of efficiently simulating small room acoustics. *J. Acoust. Soc. Am. 65,* 943-950.

Allen, J.B., McDermott, B.J., and Berkley, D.A. (1979). A method for measuring subjective perception and preference of small room reverberation. Unpublished manuscript.

Berggren, S., Berkley, D.A., McDermott, B.J. (1979). Dichotic perception of small room reverberation. Unpublished manuscript.

Berkley, D.A., Curtis, T.H., and Allen, J.B. (1973). Effects on speech perception of modifying the impulse response of a small room. *J. Acoust. Soc. Am. 53,* 30A.

Jetzt, J.J. (1979). Critical distance measurement of rooms from the sound energy spectral response. *J. Acoust. Soc. Am. 65,* 204-211.

Kruskal, J.B., and Wish, M. (1978). *Multidimensional scaling.* Sage University Paper series on Quantitative Applications in the Social Sciences, 07-011. Beverly Hills and London: Sage Publications.

Mintzer, D. (1950). Transient sounds in rooms. *J. Acoust. Soc. Am. 22,* 341-352.

Morse, P.M., and Ingard, K.U. (1968). Sound waves in ducts and rooms. *Theoretical Acoustics,* 467-608. New York: McGraw-Hill.

Oppenheim, A.V., and Schafer, R.W. (1975). *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall.