ELECTRONIC SIMULATION OF AUDITORIUM ACOUSTICS

by

Jeffrey Borish

CCRMA
DEPARTMENT OF MUSIC
Stanford University
Stanford, California 94305

ELECTRONIC SIMULATION OF AUDITORIUM ACOUSTICS

by

Jeffrey Borish

Developing a model for electronically simulating a concert hall comprises two phases, synthesis and analysis. The model proposed in the synthesis phase is based on the familiar image model, but an extension allows it to deal with complex geometries. The inputs to the model are the geometry of the concert hall, the positions of the source and listener, and the acoustical properties of the surfaces. The results can be reduced to the directional impulse response and convolved with an audio signal to create audible simulations, and are also useful in analytical studies. For example, the extended image model can demonstrate that rectangular concert halls have a fundamental advantage over the fan shape, and that another shape never before applied, the reverse fan, would be better still. The image model can be further refined to deal with diffusion, angle- and frequence- dependent reflectance, and nonisotropic sources.

Analysis of this concert hall model begins with suitable measurements of existing concert halls. Measuring with noise instead of an impulse surmounts dynamic range limitations. The desired impulse response is extracted from the response by crosscorrelating it with the input. Using a maximal-length sequence as the excitation makes it possible to perform the crosscorrelation very efficiently. Because the pseudo noise is a binary waveform, the crosscorrelation requires only additions and subtractions. Furthermore, the fast Hadamard transform can be used to minimize the number of computations.

To evaluate the model, these measurements can be compared to the computed impulse for the same hall, but the comparison must account for subjective response. Rather than attempting to objectify subjective response, a better approach is to base the comparison on a subjective evaluation. Binaural presentations preserve the directional characteristics while minimizing the practical difficulties of measuring the concert hall. Computing the binaural impulse response requires an extensive set of measurements of the head-related transfer function.

Domestic audio reproduction is a less demanding application requiring only two additional properly-positioned speakers for an adequate presentation. The signal processing is minimized by taking advantage of the reverberation present in stereo recordings. The properties of auditory perception permit additional simplifications.

DEDICATION

To my wife Claudia for her love and support.
To my parents for instilling a life-long love for the aesthetic appreciation of sound.
And to my brother, now only half a pair o' docs.

CURRICULUM VITAE

Mr. Borish received the S.B. degree in Electrical Engineering from M.I.T. in 1974, and the M.S. degree from Stanford University in 1975. He worked for several years, first at Fairchild Semiconductor designing ICs for audio applications and then at Sound Technology designing digital signal processors, before returning to Stanford to pursue the Ph.D. degree. Concurrent with his academic work, he is cofounder of Signal Research Laboratory, which specializes in new applications of digital signal processing to audio problems, and he consults for audio companies.

An avid music buff, Mr. Borish is an amateur pianist and a frequent concert-goer. He is a member of Tau Beta Pi, Eta Kappa Nu, AES, IEEE, and ASA.

TABLE OF CONTENTS

ix

CHAPTER 1

INTRODUCTION

1.0 BACKGROUND

There is a famous anecdote about Philharmonic Hall. From the moment the new concert hall at Lincoln Center opened in 1962, listeners complained about its acoustical problems. Desperate for a solution, the New York Philharmonic turned to George Szell for a recommendation. Szell, who is remembered not only for his musical genius but also for being bluntly outspoken, responded with the poisonous prescription [NEW YORK TIMES-70], "Tear the place down and start again. The hall is an insult to music." The owners of the hall opted for less drastic remedies. But after five attempts to improve the acoustics of the hall failed, Szell's prescription was finally taken. Today the hall that rose from the ruins ranks among the best in the country, but it was obtained only through unusual perseverance, enormous expense, and a fourteen-year delay.

Fiascoes like Philharmonic Hall lead laymen to infer that architectural acoustics is still a black art. Yet the opening of the century saw the opening of a concert hall that was claimed to be the first designed according to scientific principles. Today, Boston's Symphony Hall is

considered one of the three best in the world. Following this signal success, much more scientific research was performed, giving the designers of Philharmonic Hall the confidence to boast that it too would be a great hall. Scientific principles do not grow stale with age, so Philharmonic Hall should have been at least as good as Symphony Hall. That it was not lends credence to the uninformed view of unimpressed listeners.

To be sure, concert hall design is still an imprecise science. Despite our growing technical sophistication, outstanding acoustical characteristics elude many designs. There are a number of reasons it has proven difficult to repeat the success of Symphony Hall. Economic factors make it impossible to simply copy Symphony Hall. There is an ever-intensifying pressure to increase the seating capacity of concert halls. Symphony Hall seats just over 2600 people. Typical of a new style of concert hall design, Louise M. Davies Hall in San Francisco, which opened in 1980, seats over 3000. To maximize their utility, many halls attempt to satisfy divergent criteria. It is not unusual today for a new hall to be used for dance and theater as well as symphonic music. It is probably unreasonable to expect one hall to serve many uses well.

Another consideration is that, despite its high regard, Symphony Hall has limitations that modern designs should try to remedy. The sound deep under the first balcony is poor. The view from many locations is partially obstructed. Also, the seats are very uncomfortable, and would not be considered acceptable by modern-day standards.

Yet another factor is that architectural acousticians either do not want to or are not able to limit architects to incremental changes from proven designs. Radical designs might seem like dramatic jumps into the future, but when the designs fail—and they usually do—it is nearly impossible to learn from the failure. Ideally new concert halls would differ in one parameter— or at most a few—so that subjective differences could be attributed to particular architectural features, allowing the merit of these features to be evaluated. Nevertheless, it seems that most people, professional and layman alike, reserve their highest regard for designs that are most innovative despite the fact that such an approach could not be considered good science.

Finally, the continuing difficulties in designing concert halls force one to conclude that there must be significant factors of which architectural acousticians simply have not been aware. Auditorium design is an extremely complex problem not only because the physics is very

complicated, but also because the evaluation involves subjective response. Analytic principles can be applied to anticipate subjective response, but first psychoacoustic studies must identify the subjective correlates of architectural features and reduce them to quantitative terms. In retrospect it seems clear that luck contributed to Sabine's success with Symphony Hall because he could not have known all the parameters that must be controlled. The inadequacies in our understanding became apparent only with the completion of ambitious designs that attempted to satisfy more stringent visual and economic constraints. The advances that have been made in applying new materials and geometries to satisfy these constraints have outpaced advances in our understanding of the psychoacoustics of performing spaces.

1.1 ELECTRONIC REVERBERATION

Electronic simulation of concert hall acoustics is a technique that might make it possible to avoid future fiascoes. Sophisticated computer modeling and digital signal processing would allow architectural acousticians to audition concert halls before they are constructed. A computer could compute the acoustical properties of the hall from a description of its physical characteristics. The results of the analysis could then be used to process a dry audio signal, creating the impression that the sound had been emitted in a physical realization of the hall. The techniques would also be useful to preaudition proposed modifications to an existing hall. Any deficiencies in the perceived reverberation could be corrected before corrections became expensive.

Acousticians do use physical scale models of concert halls for studying the acoustics of a hall being designed or modified [BARRON-83][JORDAN-75][BURD-75], but their effectiveness is limited. Not only is building scale models expensive and time-consuming, but it is difficult to scale measurements back to reality. Scaling is usually performed by recording the response on a tape recorder operating at high speed and reducing the speed for playback. Scaling the geometry down requires that the range of frequencies be scaled up by the same amount. Even an 8:1 scale—the largest used—requires frequencies up to 160 kHz to cover the audible spectrum. Finding material that accurately scales the absorptive properties at audible frequencies to this higher range is difficult. Furthermore, air is much more absorptive at high frequencies. To counteract this effect the air in the model can be dried or nitrogen can be substituted [DAY-75],

but either technique is cumbersome. Transducer design is another stumbling block, not only because reaching the high frequencies involved is difficult, but because the transducers become highly directional. These problems, which are even more serious for more common scale factors of 40:1 or more, obviously do not exist in electronic simulation.

Although preauditioning concert hall designs is one important application of electronic reverberation, there are several others as well. A second is adding spatial effects to otherwise dry musical performances. This application is important to the recording industry because it simplifies the recording process. The musical selection can be recorded in an environment and in a manner best suited to the recording engineer. Then the desired level of reverberance is obtained by mixing artificial reverberation (which can be but is not necessarily generated electronically). This approach permits the recording engineers to devote their attention to getting the best sound for each instrument, obtaining the proper balance between all the performers, creating the desired sense of position, and overcoming background noise without simultaneously having to worry about providing a pleasing reverberation. Synthetic reverberation is also important to composers of electronic music when it is their desire to create the impression that the synthesized sounds are being performed in a closed space.

The third application is adding a reverberant effect to domestic audio reproduction. Accurate reproduction of reverberation requires that sounds reach the listener from directions outside the range stereo is capable of presenting. Rather than developing a new recording standard whose attraction would inevitably be limited by commercial considerations, another valid approach is to synthesize reverberant sounds during playback that model the behavior of performing spaces. This approach has the added advantage of allowing the listener to control the parameters of the reverberation according to his tastes; he is no longer constrained by the whims of the recording engineer.

A final area of interest for synthetic reverberation is in studying the perception of reverberation. Traditionally, these studies have been carried out by traveling from one hall to another to listen to musical performances and then trying to associate architectural differences with the perceived acoustical differences. Such studies are vitiated by several factors. The first is the confusion produced by differences in the performed music. The second is the difficulty of preserving an acoustical memory during the inevitable time lapse in traveling from one hall to

another. There is also the problem of isolating the cause of the perceived differences because halls generally differ in not one but many regards. Attempts to obtain quantitative measures of the differences between halls have frequently been confounded by the difficulty of obtaining measurements with sufficient resolution to be meaningful. In addition to minimizing these problems, studies based on synthetic reverberation provide an another advantage: parameters can be manipulated that are not easily controlled in existing halls. Electronic simulation allows us to ask questions such as "what is the subjective effect of repositioning a wall," or "what happens when absorbing material is introduced." Such changes could be as simple as turning a knob in a reverberation synthesizer, but could require huge expenditures when actual structures must be changed. These psychoacoustic studies would be valuable not only for increasing our understanding of auditory perception, but would also have practical application when the results are formulated in terms of design criteria for concert halls.

1.2 SYNTHESIZING A CONCERT HALL MODEL

This thesis describes a methodology for developing an electronic model for concert halls, and also some preliminary results of applying this methodology. The development of any model comprises two phases, synthesis and analysis. In the synthesis phase, the researcher must use his understanding of basic principles and his intuition to propose a new model or to refine an existing one. This model is then tested in the analysis phase by comparing it with reality. Discrepancies indicate flaws in the model, and, in some cases, suggest ways in which the model can be improved.

The model proposed in the synthesis phase (Ch. 2) is based on the familiar image model. The inputs to the model are the geometry of the concert hall, the positions of the source and listener, and the acoustical properties of the surfaces. The model finds the paths along which sound travels to the listener. Quantitative studies of the acoustical properties of concert halls often begin by determining the sound paths. For example, we show how the image model can be used to explain why rectangular concert halls sound better than fan-shaped halls. The image model also reveals that another shape never before applied in practice, the reverse fan, would probably be even better than the rectangular hall.

Knowing the sound paths also makes it possible to determine the impulse response of the concert hall for the given source and listener positions. Then audible simulations can be created by convolving this impulse response with an audio signal. One complication in creating these audible simulations is that it is necessary to convey the directions from which the various reflections should reach the listener. Basing all audible presentations on a binaural signal provides the desired directional cues with a practical system. An extensive set of measurements of the directional impulse response of a subject's ears is used as the basis for the required model of binaural hearing.

1.2.1   Previous work in computer modeling

Mathematical modeling has often been used to analyze the acoustical properties of enclosed spaces, but little of this work has been directed specifically towards creating audible simulations. Rather, modeling has been used in quantitative studies of particular aspects of concert hall acoustics. For example, Santon [SANTON-76] and Plomp et al. [PLOMP-80] considered the use of the image model for predicting the intelligibility of speech in rooms. Allen and Berkley [ALLEN-79] used the image model for a variety of perceptual studies of reverberation. A series of works [BAXA-78][BAXA-80][HULBERT-82] applied quantitative criteria to rate the acoustical quality of the modeled hall and to attempt to optimize its geometric and acoustical characteristics. Mathematical models have also been used for theoretical studies of various properties of concert halls, including spatial impression [BARRON-74], reverberant decay [HIRATA-79][SCHROEDER-70][WAYMAN-77], and distribution of sound pressure level [GIBBS-72].

Many previous studies used ray tracing rather than an image model. Ray tracing tends to be easier to program, particularly when dealing with complex geometries. Although both approaches are based on the same assumptions of geometrical acoustics, they mechanize the analysis differently. In ray tracing, the source is assumed to emit a finite number of sound rays in many directions. These rays are extended by linear extrapolation and reflection until they reach a zone of space surrounding the listener. Unfortunately, ray tracing will not find all the sound paths whose length is less than a given amount. This limitation is discussed in greater detail in Ch. 2.

Those studies that did apply the image model often restricted consideration to perfectly rectangular rooms. Computing the image positions in rectangular rooms is trivial, so these programs are simpler to write, easier to use, and more efficient to execute. But real concert halls are never perfectly rectangular, so accurate modeling requires that the image model be generalized. The extension of the image model presented in Ch. 2 makes it possible to deal with polyhedra having any number of sides. Reentrant angles are permitted as well, making it possible to model obstructing surfaces such as balconies. These generalizations surmount the most obvious limitation of the image model.

The only other serious work in audible simulation of concert halls is the work of Walsh [WALSH-81][WALSH-82]. Walsh also uses a polyhedral representation for the room. Sound is traced from the source to the listener using what is essentially a ray tracing method, except that the beams have finite cross-section, allowing them to be split when they strike an edge. The sound in each path can be filtered independently to take account of the variation with frequency of the absorption characteristics. The frequency dependence is modeled in eight octave bands. The other similarity with this thesis is that Walsh also produces binaural simulations.

1.2.2 Two-phase modeling

The duration of a concert hall impulse response (as defined by the reverberation time) can be over 2 sec. For precise modeling, the entire impulse response can be computed, but the amount of computation required increases approximately exponentially, quickly becoming prohibitive. Furthermore, convolving the audio signal with such a long impulse response would also be time-consuming. The only practical solution uses an idea first proposed by Schroeder [SCHROEDER-70] to divide the reverberation into two regions. The early reflections are sparse, making it possible for them to be perceived individually to an extent. As a result, their precise timing, amplitude, and direction will usually affect our perception. The later reflections— designated reverberance—are more dense. This high echo density produces a noise-like quality, although some of the spectral characteristics of the original signal are retained. Consequently, it is primarily the statistical properties and not the detailed structure that is important. Thus, cruder synthesis techniques suffice for the reverberance, and detailed modeling can be restricted to the early reflections.

The common feature of all existing techniques for generating reverberance is recirculating loops with delay. The difficulty with such loops is that they can produce strong temporal and spectral patterns that color the sound objectionably. Most techniques to circumvent this problem have evolved from Schroeder's seminal work [SCHROEDER-61]. Schroeder modified the basic recirculating loops (comb filters) by adding part of the undelayed signal to the output producing an all-pass characteristic. But the more significant contribution was the idea of keeping the loop gains relatively small to minimize coloration, and to provide a long decay by cascading several reverberators. Schroeder later refined this technique [SCHROEDER-62] by prefixing four parallel comb filters to introduce some of the resonances characteristic of enclosed spaces. Although computationally efficient, the technique is inaccurate in that it does not duplicate the acoustic events in a concert hall. Furthermore, it suffers from some practical limitations. Changing the feedback gains and delays can easily lead to colored reverberation and a ragged decay, making it difficult to adjust the reverberation time. Also, the decay rings, and the echo density builds up too slowly [MOORER-79].

Moorer [MOORER-79] refined Schroeder's technique by introducing low-pass filters in the recirculation paths. The low-pass filters increase the realism by shortening the decay at higher frequencies. Also, the response sounds smoother because the low-pass filters smear the echoes produced by the unit reverberators. Moorer favors a system with six comb filters followed by a single all-pass.

A radically different approach was invented by Moore [MOORE-81]. Rather than a combination of unit reverberators, Moore uses a multiply-tapped delay line with recirculation. To eliminate the regular temporal patterns, he slowly varies the delay times of the taps. Continuously changing a tap delay will frequency modulate the signal, but slow variation minimizes the audibility of the modulation. Also, the rates and direction of delay change—and therefore to the pitches—of different taps are not synchronized, so that the corresponding pitch variations tend to be less obvious in the final sum.

Although these techniques have limitations, they are adequate for synthesizing reverberance, so this thesis will concentrate on the problem of accurately modeling the early reflections.

1.3 ANALYZING THE MODEL

Evaluating the model requires a standard for comparison. The standard is provided by modeling an existing hall, and comparing the results with measurements of the hall. Measuring concert halls is complicated by the difficulty of obtaining adequate dynamic range. Impulsive excitations are often used to measure the impulse response, but they have several practical difficulties—principally the fact that to obtain adequate temporal resolution, the duration of the impulse must be very short, making it difficult to inject enough energy into the concert hall to overcome the ambient noise.

A different approach is to excite the concert hall with noise. Being a sustained signal, noise injects sufficient energy to overcome the noise of the system. The impulse response can be extracted from the measurement by crosscorrelating it with the input. Ch. 3 discusses a method based on pseudorandom noise. Pseudorandom noise has the advantage of being a deterministic signal, so that the excitation does not contribute randomness to the result. A binary sequence offers the additional advantage of simplifying the crosscorrelation: assigning the values ± 1 to the two binary levels eliminates multiplications. Finally, basing the binary pseudo noise on a maximal-length sequence provides additional economies because there is an efficient algorithm that reduces the number of additions and subtractions to the order of $n \log_2 n$, where n is the number of samples in the measurement. Efficiency is a paramount concern because the concert hall impulse responses can be very long.

How these measurements are used to evaluate the model is discussed in Ch. 4. Evaluating a concert hall model is complicated by the need to take account of subjective response. Rather than attempting to objectify subjective response, the approach suggested in this thesis uses a subjective evaluation. Subjective evaluation is also complicated by the need to account for the directional characteristics of the reverberation. One simple way of preserving directional cues is to employ the binaural impulse response of the concert hall. Requiring that only two channels be measured minimizes the inconvenience associated with making measurements at remote locations. Thus, binaural presentation can be conveniently applied to both the computed and measured impulse responses.

By repeatedly using the results of the analysis for successive refinements of the model, the

development of a concert hall model becomes an adaptive process. Refining must rely heavily on intuition and insight because subjective evaluation does not directly indicate what improvements are needed. Only those physical processes that are perceptually significant should be modeled as enhancements inevitably burden the computation. Some weaknesses worth exploring are discussed in the final chapter along with possible remedies.

1.4 OVERVIEW

The concrete results presented in this thesis are a contribution to an improved concert hall model, but perhaps more important is the method that is proposed for developing and improving the model. This method provides a foundation for studies that previously were conducted desultorily. This emphasis on methodology rather than a method for modeling a concert hall offers greater freedom to candidly discuss limitations and informally propose palliatives. These suggestions could be the basis for ensuing studies. Even though the model is still crepuscular, the methodology is sound.

The methodology also provides the structure for the thesis. The next chapter deals with the synthesis phase—defining a model. The model that is proposed is based on the familiar image model, but an extension allows it to deal with complex geometries. Analysis of this model begins with suitable measurements. Ch. 3 discusses a method based on pseudorandom noise that not only provides adequate dynamic range, it does so with a minimum of computation. How these measurements are used to evaluate the model is discussed in Ch. 4. Rather than attempting to objectify subjective response, the approach adopted in this thesis uses a subjective evaluation. Subjective evaluation is also complicated because of the need to account for the directional characteristics of the reverberation. The solution is to base the presentations on binaural hearing. The penultimate chapter discusses simplifications that make it possible to meet stringent cost constraints while still providing a natural-sounding ambience that is a worthwhile enhancement of domestic audio reproduction. The thesis concludes with a chapter suggesting refinements to account for physical processes currently ignored.

CHAPTER 2

THE IMAGE MODEL

2.0 INTRODUCTION

The image model is a technique that is widely known for analyzing the acoustical properties of a space that includes reflecting boundaries [KNUDSEN-78][KUTTRUFF-73]. The method of images makes it possible to represent a boundary-value problem in terms of an equivalent problem involving multiple sources, but no boundaries. Implementation on a digital computer has led to some new applications for the image model. Gibbs and Jones [GIBBS-72] used the image model to estimate the total sound intensity in a concert hall. More recently, Allen and Berkley [ALLEN-79] presented a method for creating synthetic reverberation based on the image model. Their efforts were directed specifically toward developing a method that was efficient and easy to use. To accomplish this goal, they restricted the model to rectangular enclosures. Although attempts have been made to generalize to more complex geometries, usually some restrictions have remained. Barron [BARRON-74] limited his model to two reflections and retained some geometrical restrictions. Baxa [BAXA-76] was limited by core size to five reflections. His algorithm is similar to ours, but our technique is able, even with modest memory sizes, to compute the positions of images up to essentially any number of

reflections. Although simple models have the undeniable advantage of being comparatively simple to apply and of requiring less computation time, they are rarely representative of situations one encounters in reality. Most concert halls are very irregularly shaped, so when one is interested in modeling their characteristics more faithfully, it is important to have an algorithm that is sufficiently general.

Any reasonable assumptions for the shape of a concert hall will always involve simplifications at some scale. Boston's Symphony Hall is one that is generally regarded as being rectangular, but of course its ceiling is coffered, the walls have alcoves with statues, and the floor has seats with people. It would be impossible to account for every door knob and every light bulb in the hall. But even when consideration is restricted to the macroscopic geometry, it would be useful to generalize the image model. Symphony Hall has angled reflecting surfaces around the stage that probably have a significant effect on the sound reaching frontal locations. Other halls deviate even more dramatically from the ideal rectangular shape, such as fan- or horseshoe-shaped halls. Also, most halls have balconies that obstruct a significant portion of the sound directed toward seats underneath them. This chapter shows how the image model can be generalized to deal with all of these situations. The only constraint on the degree of complexity is the amount of time required for the computations.

2.1 THE ALGORITHM

In this section we describe the algorithm for computing the image positions in a polyhedron with an arbitrary shape. The basic principle of the image model is that a path involving reflections can be represented by a straight line path connecting the listener to a corresponding virtual source (VS). When this idea is applied to a rectangular room, a regular lattice of VSs results (Fig. 2.1). Because of the regularity of this lattice, the calculation of the VS positions is trivial [GIBBS-72][ALLEN-79]. For irregular shapes it is more difficult to find the positions. Moreover, it is necessary to consider in irregular shapes that a VS might not be "visible" from every position in the room (Fig. 2.2). Applying the image model to irregular shapes not only involves a more complicated procedure for finding the positions of the VSs, but a visibility test must be performed to determine which VSs are visible.
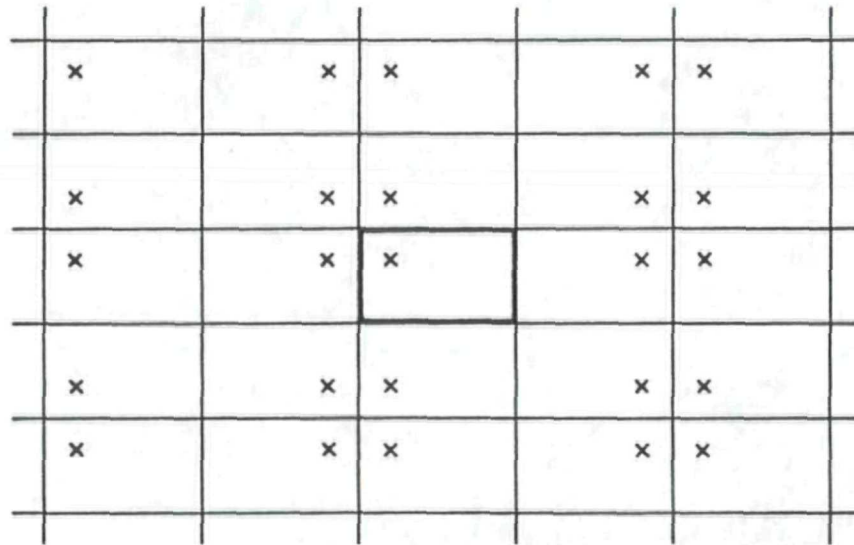
Fig. 2.1 For a rectangular room, the lattice of VSs is very regular making it easy to compute their coordinates. The accentuated box in the center represents the real room, and surrounding it are the virtual rooms each containing a single VS.

2.1.1 Finding the image point

Given a reflecting surface with an arbitrary orientation and a source point, we need to find an expression for the position of the image point. The position and orientation of the plane of the reflecting surface are described by two parameters, the unit normal to the plane $\hat{n}$, and the distance from the origin to the plane p. Referring to Fig. 2.3, the image point can be found by traveling from the real point a distance $2d$ in the direction of the planar normal. d, the distance from the point to the plane, is given by

$$d = p - \overline{P} \cdot \hat{n}, \qquad (2.1)$$
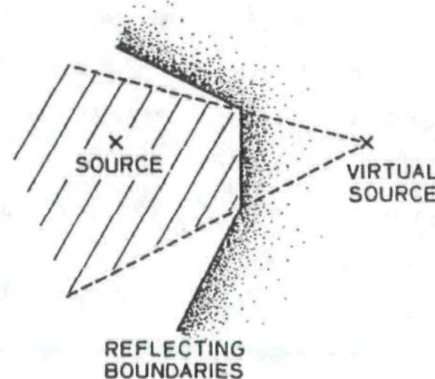
Fig. 2.2 The listener must be in the shaded region for the virtual source to be visible.
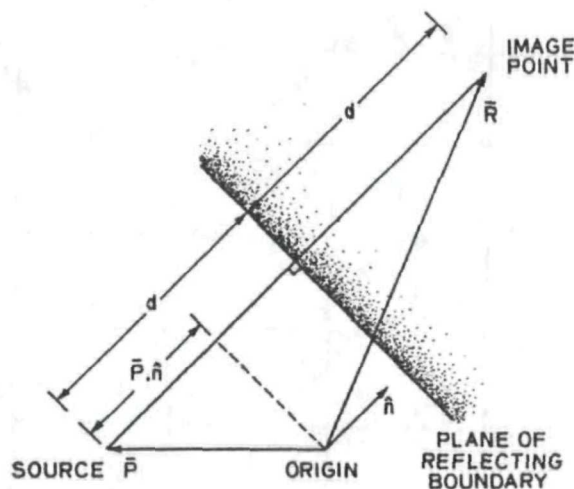
Fig. 2.3 Diagram showing how to determine the position of the image point. Its coordinates are found by traveling a distance of $2d$ in the direction of the normal.

so that $\overline{R}$, the position vector of the image point, is

$$\overline{R} = \overline{P} + 2d\hat{n}. \tag{2.2}$$

Finding the image point requires seven additions and six multiplications, compared to only one addition for the rectangular lattice of Fig. 2.1. The general algorithm will clearly require a great deal more computation.

2.1.2   Generating the lattice of virtual sources

Now that we have a means for finding the position of the image of a point in a reflecting boundary, we need a way of applying this operation systematically to assure that every VS is found. The basic concept of the algorithm is very simple: reflect every VS across every reflecting surface. This algorithm is recursive because reflecting a VS across a boundary will produce a new VS which in turn must be reflected across every boundary. Calculating the VS positions can be represented by a tree. The nodes in the tree represent the VSs, and the branches are formed when a VS is reflected across a boundary. The order of a VS indicates the number of reflections required to create it. The VSs created by reflecting across the sides of the polyhedron are called the progeny VSs.

Each VS created must be qualified by three criteria. The first is called validity. By considering the boundaries of the polyhedron to be mirrors with the reflective sides facing into
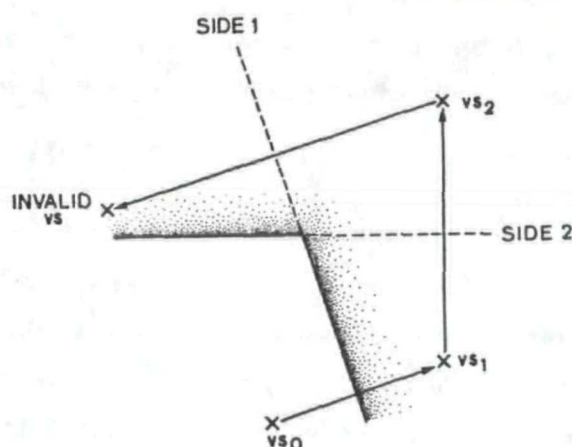
Fig. 2.4 When vs_1 is reflected across side 1, or vs_2 is reflected across side 2, invalid VSs are created. These cases could be avoided by not reflecting a VS across the side used to create it. However, it is also possible for an invalid VS to be created by routes not easily anticipated. These cases must be detected by an explicit test.

the room, an invalid VS can be defined to be one created by reflecting across the nonreflective side of the boundary. The most straightforward way this problem can arise is to reflect a VS back across the same side used to create it. This particular situation could be easily avoided by remembering which side was used to create a VS, and then skipping this side when the VS is propagated. However, it is also possible for invalid reflections to arise from more complicated routes, such as the one shown in Fig. 2.4. One solution to this problem is to test explicitly for validity each time a VS is created. Invalid reflections can be detected very easily as part of the computation of the position of the progeny VS. If the distance computed in Eq. (2.1) is negative, then the reflection is invalid. Otherwise, qualification continues to the second step.

The second criterion a VS must satisfy is proximity. The user of the program inputs a distance. VSs farther from the listener than this distance are discarded and the line of descent is terminated. It is impossible for its progeny to be closer to the listener. The proximity criterion is what makes it possible for the algorithm to terminate. Although it would also be possible to terminate the propagation on the basis of another criterion, *e.g.* by finding all VSs up to some order, proximity is the one that assures that every sound path shorter than a selected maximum will be found. This completeness is important for accurately characterizing the behavior of a room within a certain period.

The third criterion a new VS must satisfy is visibility. The visibility test is the most involved by far. In general, it must be performed in several steps. The first step is always to

ascertain that the progeny VS is visible to the listener in the side that created it. As shown in Fig. 2.2, when obtuse angles are allowed it is possible the listener will not be able to "see" the progeny VS.

To test for visibility, a line is formed joining the new VS with the listener, and the point of intersection of this line with the plane of the reflecting side is determined. If the point lies inside the boundary of the side, the VS is visible. There are many ways of testing whether the intersection is inside the boundary. One straightforward method is to form vectors from the point of intersection to each of the vertices of the side. The cross products of successive pairs of these vectors are always a vector orthogonal to the plane of the side. If each of the normal vectors so calculated points to the same side of the reflecting boundary, then the point of intersection is inside the polygon. Otherwise it is outside, and the VS is not visible. Fig. 2.5 illustrates these two possibilities. The algorithm can be explained heuristically by imagining one is standing at the point of intersection looking at each vertex in turn. As long as viewing successive vertices requires rotation in the same direction, one is inside. Having to reverse one's direction of rotation indicates the point is outside.
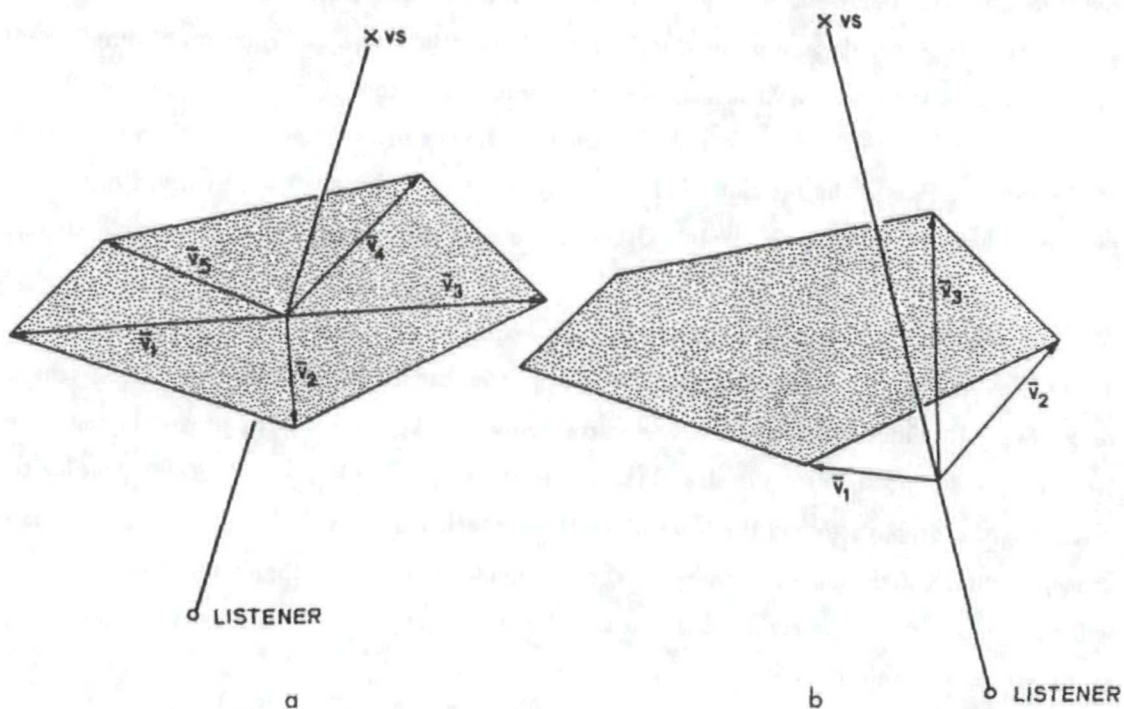
Fig. 2.5   The test to determine whether the point of intersection is inside the boundary of the side is performed by computing the crossproducts of the vectors from the point of intersection to each of the vertices in turn. In (a), all of the crossproducts point up, so the point is inside the boundary. In (b), $\bar{v}_1 \times \bar{v}_2$ points down, but $\bar{v}_2 \times \bar{v}_3$ points up, so the point is outside.

In practice, the vectors from the point of intersection to the vertices are formed as they are needed for calculating the cross product. As soon as a cross product is calculated that points in the direction opposite to previous ones, the test is terminated. Only when the point is inside the periphery is it necessary to carry the test to completion. As a result, tests that fail are on the average half as long as tests that pass. The cross product requires 6 multiplications, so the number of multiplications required by the algorithm is no more than $6v$, where v is the number of vertices on the side.

For first-order reflections this test is sufficient. Higher-order reflections require additional visibility tests. Consider the second-order reflection in Fig. 2.6. By decomposing the path from the VS to the listener into the actual path the sound travels, it is clear that if side 2 extends to point A the path will exist, but if it extends only to point B the path will not exist. An additional visibility test is necessary to validate segment a-b. This test can be performed by looking for vs_1 in side 2 from the position that is the mirror image of the real listener position. Paths involving more reflections require higher-order virtual listeners that look for the corresponding
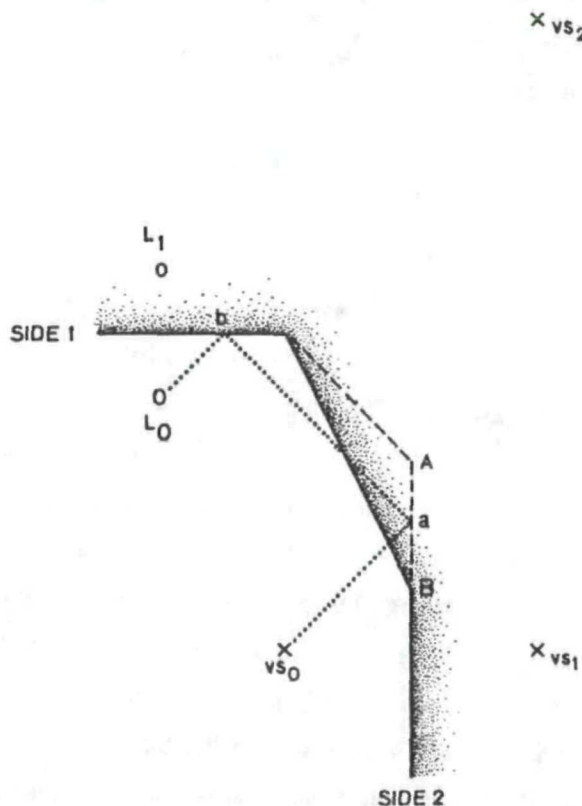
Fig. 2.6   Although vs_2 is visible in side 1 from the listener position, vs_2 still will not be visible if side 2 extends only to point B. Validating segment a–b of the sound path requires that virtual listener L_1 look for vs_1 in side 2. L_1 will not see vs_1 if side 2 ends at B.

VSs in order to validate every segment of the real sound path.

Only visible VSs contribute to the sound that reaches the listener. However, invisible VSs are retained because their progeny might still be visible. A VS is discarded only when it fails either of the first two tests, proximity or validity.

The number of VSs at each level of the tree increases rapidly. The number of VSs at level k is approximately N^k, where N is the number of sides in the polyhedron. This expression is only approximate because not every VS survives the first two qualifying tests. Because the number of VSs increases so rapidly, remembering all of them at the same time would require an excessive amount of storage. The visible VSs usually represent a small part of the complete set of VSs. The invisible VSs need to be remembered only temporarily until all their descendants have been found. By traversing the tree using a preorder sequence [KNUTH-73] rather than on a level-to-level basis, only the VSs in a direct line of descent need to be stored. Simulations rarely exceed eighth order, so the memory requirements can be vastly reduced. When a VS is visible, it is placed in a separate table of visible VSs. Otherwise it is forgotten as soon as it is no longer needed to propagate the tree.

2.1.3   Obstructions

All the discussion to this point has assumed that the polyhedron is convex. Allowing reentrant angles makes it possible for a sound path to be obstructed. A side with the potential of obstructing a sound is designated obstructive. Most, if not all, concert halls have obstructive sides. The obstructor might be as innocuous as a wall that splays outward, or it could be a balcony (Fig. 2.7). Because obstructive surfaces are so common, it is important that the image model be able to deal with them. Testing for obstructions must be performed for each segment of the path, just as for visibility testing. The point of intersection of the line connecting the test VS and test virtual listener with the plane of each potentially obstructive side is calculated. If the intersection is inside the boundary of the side, then the view is obstructed. Because obstruction testing adds a considerable burden to the visibility test, obstructive sides are tagged so that the test will not be applied to sides that cannot obstruct the sound. Obstructive sides are treated normally as far as reflections are concerned. Thus, a balcony can also produce reflections, as shown in the figure.
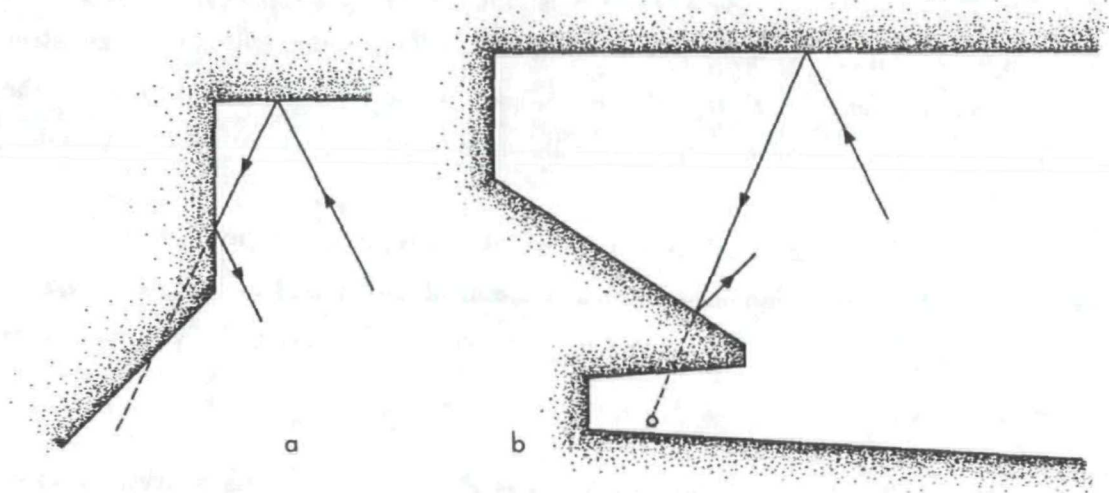
Fig. 2.7 In (a), an obstruction is formed by a splay in a side wall. (b) illustrates the familiar case of a balcony obstructing a sound.

Testing whether the point of intersection is inside the boundary of the side is more complicated when reentrant angles are allowed. As seen in Fig. 2.8, the direction of rotation can reverse even when the point is inside the boundary. One solution is to sum the angles of rotation in moving from one vertex to the next. If the total is 2π radians the point is inside. If it is 0 the point is outside. Unfortunately, this remedy significantly increases the complexity of the visibility testing because calculating the angles involves inverse trigonometric functions and also because the test must be carried to completion even when the point is outside.

With some shapes—the side in Fig. 2.8 is an example—bypassing the reentrant vertex
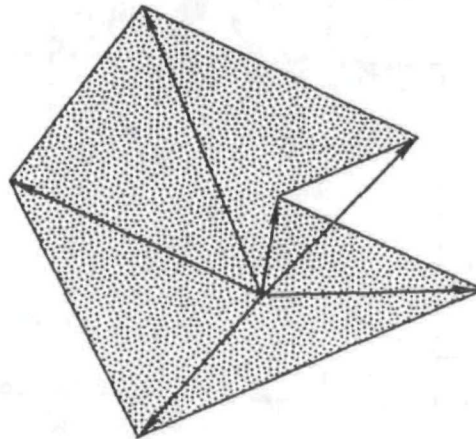
Fig. 2.8 When reentrant angles are allowed, the direction of rotation can reverse even when the point is inside the boundary. Therefore the test must measure the total rotation. 2π radians indicates the point is inside; 0 radians indicates it is outside.

produces a concave polygon, allowing the simple visibility test to be applied. Of course, this distortion will confuse the visibility test when the point of intersection falls in the nonexistent triangular region. But to reach the triangle, the path will also have to penetrate one of the corresponding obstructive sides, so the path will still be rejected, albeit for a different reason.

Yet another solution is to subdivide the side into coplanar regions, each one of which is convex (Fig. 2.9). Of course doing so increases the number of VSs that must be calculated, but the additional computation is partially offset by the savings in being able to use a simpler visibility test.

The complexity of the visibility test increases as the interior angles of the polyhedron increase. An acute polyhedron (one whose interior angles are no greater than 90°) does not require visibility testing. A concave polyhedron (one whose interior angles are no greater than 180°) does not require obstruction testing. The worst case is a polyhedron with reentrant angles, which requires both visibility and obstruction testing.

The extended algorithm is summarized by the notional computer program in Fig. 2.10. The procedure that generates the tree, called PROPAGATE, is recursive. The program is initiated by calling PROPAGATE with the coordinates of the real source. The function of the subroutines should be clear from the preceding discussion. Whenever a valid and proximate VS is found, the procedure invokes itself with the new VS. Consequently, the tree traversal proceeds to the tips before completing the propagation at any one level.
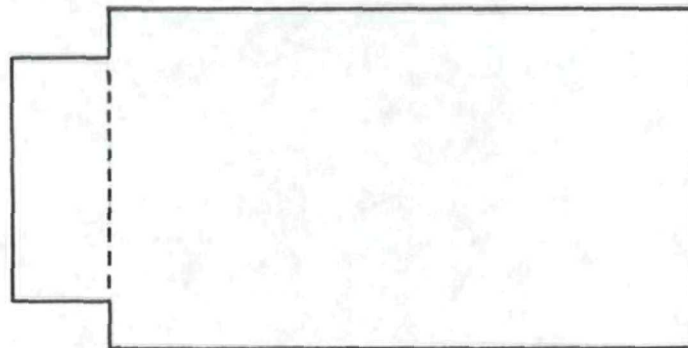
Fig. 2.9   The simpler visibility test can be applied when it is possible to subdivide a convex side into coplanar concave regions. This technique produces more VSs, but the additional computation is offset by a simpler visibility test.

```
RECURSIVE PROCEDURE propagate(vs);

BEGIN "propagate"

    FOR side ← 1 STEP 1 UNTIL nsides DO

        BEGIN "reflect vs across each side"

        progeny_vs ← reflection(vs, side);

        IF NOT valid(progeny_vs) THEN CONTINUE "reflect vs across each side";

        IF NOT proximate(progeny_vs) THEN CONTINUE "reflect vs across each side";

        IF visible(progeny_vs) THEN

            IF NOT obstructed(progeny_vs) THEN remember(progeny_vs);

        propagate(progeny_vs);

        END "reflect vs across each side";

    END "propagate";
```

Fig. 2.10 The notional computer program for the extended image model. Each valid and proximate VS is recursively propagated. Only visible ones are remembered.

2.2   USING THE RESULTS

In this section we show how the extended image model can be used to analyze and simulate concert hall geometries.

2.2.1   Displaying the results

Because of the three-dimensional character of the data produced by the image model, displaying the results in a meaningful manner requires some consideration. The most straightforward presentation is a perspective view of the three-dimensional scene made up of the concert hall, the source, and the surrounding VSs (Fig. 2.11a). Also in this display are three imaginary walls upon which are drawn the projections of the three-dimensional scene along each of the coordinate axes. The user is able to change his vantage. It is also possible to view only the projections on the imaginary walls.

Another type of display is illustrated in Fig. 2.11b. This display shows the positions of the VSs in the "angle plane." The vertical axis is the elevational angle, and the horizontal axis is the azimuthal angle. The size of the X indicating the position of a VS can be used to suggest either its distance from the listener or the amplitude of its sound. This type of plot is very useful for displaying the diversity of directions sounds emanate from. Some concert hall geometries tend to compress the VSs into smaller regions; others scatter them about more uniformly.
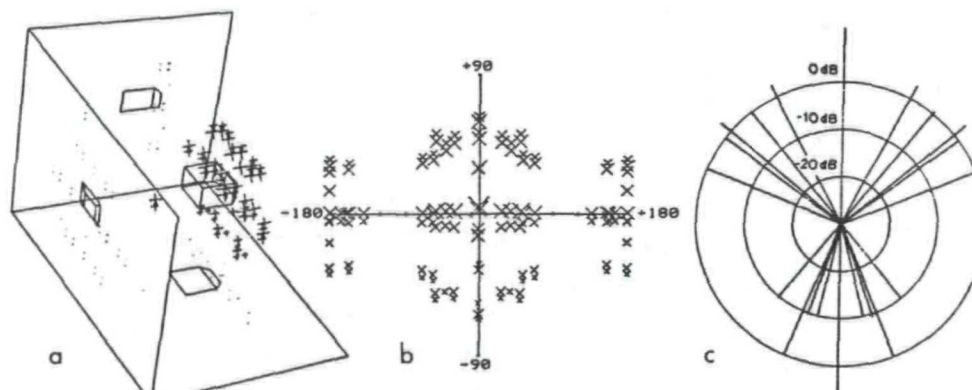
Fig. 2.11   Three different techniques for displaying the positions of the VSs. (a) is a perspective view of the concert hall with its VSs. (b) shows the positions of the VSs in the angle plane, with the size of the $\times$ indicating the strength of the VS. Finally, (c) is a polar plot of sound power as a function of azimuth (the elevation is ignored). All the displays are based on an idealization of Symphony Hall.

A third type of display is a polar plot of sound power (Fig. 2.11c). This display gives a very clear indication of the extent sound is spread to the sides. In the polar plot as in the angle plane, 0° azimuth is defined as the direction of the source. We will find different circumstances in which each is the most appropriate presentation.

2.2.2   Calculating the directional impulse response

The primary application of the image model is creating audible simulations of concert halls. To create the impression that the sound was produced in a physical realization of the hall, a dry audio signal must be convolved with the impulse response of the hall. Reducing the VS positions to the impulse response is straightforward. The distance between the source and listener produces a delay of

$$t_0 = \frac{\|\overline{vs}_0 - \overline{l}\|}{c}, \tag{2.3}$$

where the vectors $\overline{vs}_0$ and $\overline{l}$ represent the source and listener positions, and c is the speed of sound. It is impossible for a listener to detect this delay time using only his auditory sense, so the arrival times of all the echoes are referenced to the arrival time of the direct sound:

$$\tau_i = \frac{\|\overline{vs}_i - \overline{l}\|}{c} - t_0, \quad i \geq 0, \tag{2.4}$$

where $\overline{vs}_i$ is the vector to the ith VS.

Having found the arrival times of the delayed sounds, one must next determine their amplitudes. It is conventional in the image method to consider the attenuation as arising from two factors. First, the sound pressure level depends on the reciprocal of the distance traveled because the sound, assumed to have been produced by a point source, propagates in a spherical wave. Second, sound energy is absorbed at the reflecting boundaries. Thus, the amplitude of a delayed sound relative to the direct sound is given by

$$g_i = \frac{R_0}{R_i} \prod_{j \in S} \Gamma_j, \tag{2.5}$$

where R_i is the distance of the listener from the ith VS, S is the set of sides the sound encounters, and Γ_j is the reflection coefficient of the jth side. It is also possible to include an exponential term to model the distributed absorption by the air itself [GIBBS-72].

In presenting the simulated reverberation to a listener, it is important that some means be found for conveying the directional dependence of the impulse response. The position of a VS relative to the listener indicates the direction from which its sound will reach the listener. The most precise presentation is obtained by positioning speakers around the periphery of the room in directions corresponding to the directions of the virtual sources. The angle plane plot (Fig. 2.11b) gives a clear indication of the required directions. The signal that drives each speaker must be delayed and attenuated appropriately to account for the distance between the speaker position and the position of the corresponding virtual source. Clearly, the large number of speakers required makes this approach impractical, although it is precise. Several levels of approximation are available depending on the desired degree of precision. These approximations will be discussed in later chapters.

2.2.3 Influence of concert hall geometry on subjective response

In addition to providing a means for determining the impulse response of a concert hall, the image model is also capable of providing insight into ways in which the concert hall geometry affects our subjective response. Many of the concert halls widely regarded as the best in the world are rectangularly shaped (Fig. 2.12). Other common shapes, such as the fan, have rarely been found to provide the same degree of listener satisfaction. Although it is conceivable that the differences are coincidental—having to do with the particular implementations, choice of materials, *etc.*—their consistency strongly suggests that a fundamental law is at work.
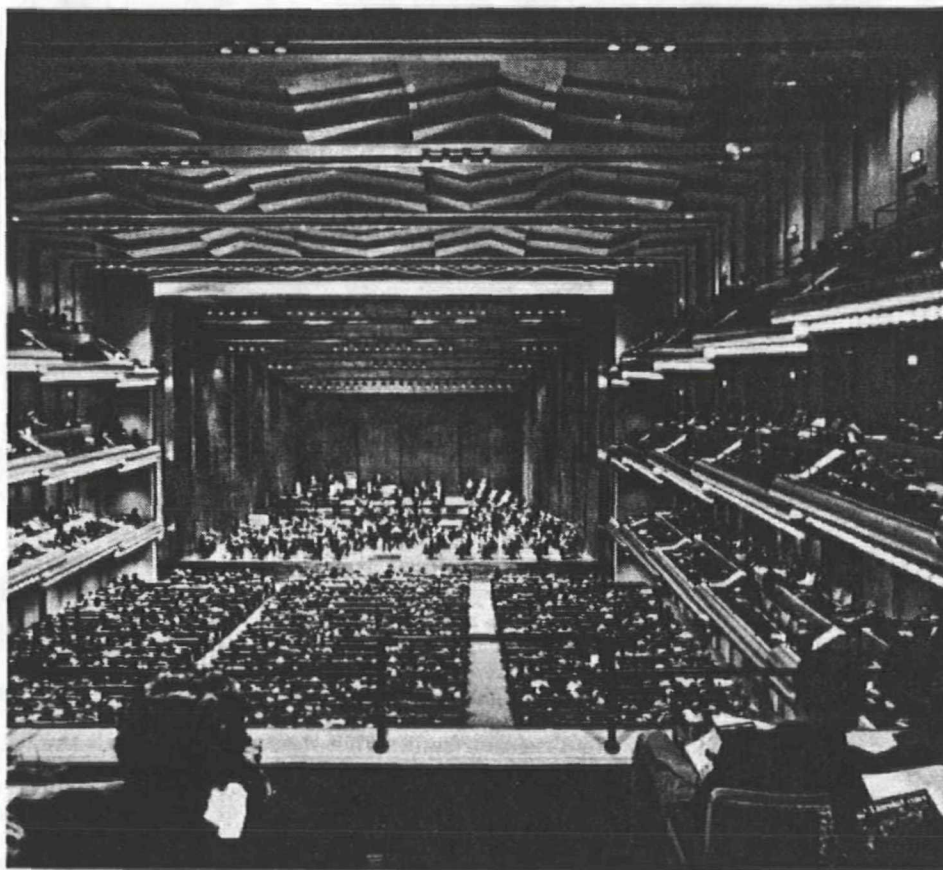
Fig. 2.12 Avery Fisher Hall is an example of a hall with a classical shoebox shape, which it shares with the best concert halls in the world. Its predecessor, Philharmonic Hall, had a basically fan shape, and was razed when repeated attempts to remedy its acoustical deficiencies failed. (Picture courtesy Lincoln Center for the Performing Arts/McGrath)

To approach this problem using the image model, displays were generated for a sequence of concert halls that vary from a fan-shaped hall, to a rectangularly-shaped hall, and finally
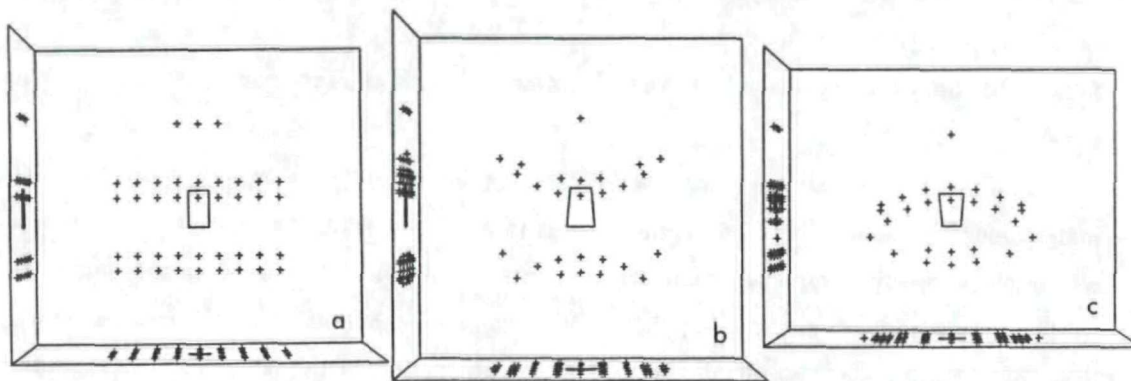
Fig. 2.13 The VSs spread to the sides along straight lines in the rectangular hall in (a). When the side walls splay outward as in (b), the VSs curve toward the front of the hall, decreasing the degree of spatial impression. By contrast, in the reverse fan (c) the rearward curvature increases spatial impression.
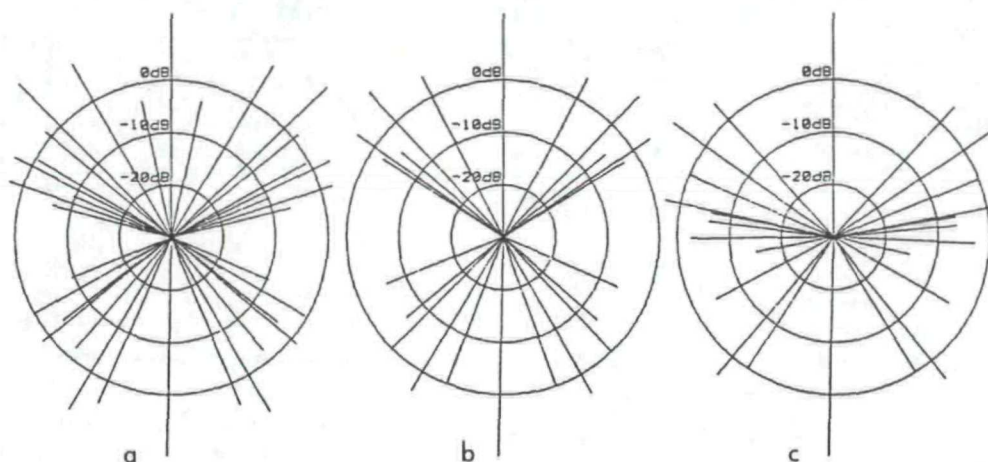
Fig. 2.14   The polar plots for the same sequence of halls from (a) rectangular to (b) fan to (c) reverse fan emphasize that the lateral sound is more centralized in the fan hall and less in the reverse fan than the rectangular hall.

to a reverse fan. Viewing the VS positions from directly above to suppress the elevational component (Fig. 2.13), we observe an interesting progression. For the rectangularly shaped hall, the VSs spread out to the side in straight lines, as expected. Because the computation of VS positions is terminated on the basis of their distance from the listener, only the beginning of the regular pattern is visible. When the side walls are splayed outward in the fan-shaped hall, the lines of VSs curve toward the front of the hall. Conversely, when the side walls splay inward in the reverse fan, the VSs curve rearward. From the listener's perspective, the fan shape compresses the sound toward the center whereas the reverse fan spreads it to the sides, a fact which is emphasized in the polar plots for the same sequence of halls (Fig. 2.14). Barron [BARRON-74] has described the subjective effect of lateral delayed sounds as spatial impression (SI). The farther from the median plane, the greater the SI. Of the three hall shapes examined, the fan-shaped hall would show the lowest degree of SI. The rectangular hall would be expected to be preferable, as is often found to be the case in practice. What is somewhat unexpected is that the sequence continues to the reverse fan which should be the best of the three. This shape has never been applied in an actual concert hall.

The reason the reverse fan-shaped hall has never been attempted in practice presumably has to do with its practical limitations. One of the economic pressures in designing new concert halls is to increase the seating capacity. The fan shape is a natural outgrowth of this desire as it provides a large number of seats with acceptable sightlines. Because the width of the front part of the concert hall is constrained by the requirements of the performing groups, the
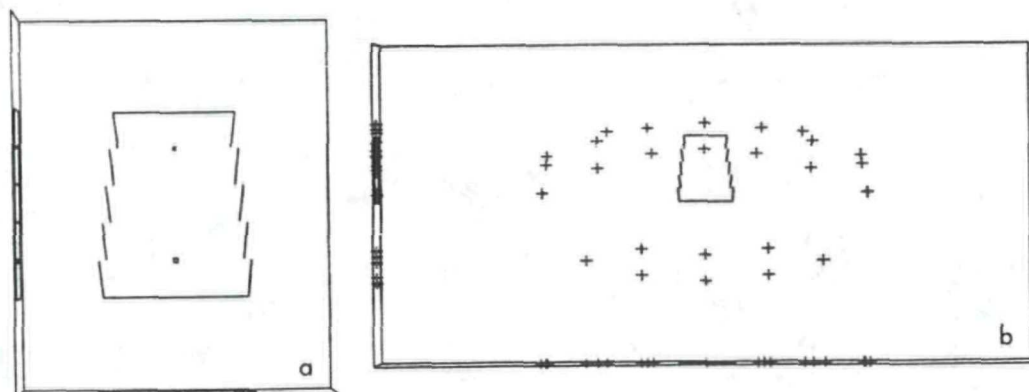
Fig. 2.15   The segmented fan combines the acoustical characteristics of the reverse fan with the practical advantages of the fan by positioning the segments according to the fan but orienting them according to the reverse fan. As shown in (b), the VS positions show the same rearward curvature as in the reverse fan, so the two should have comparable spatial impression.

reverse fan would sacrifice seating capacity. Furthermore, the reverse splay of the side walls would probably create some problems in the appearance of the hall.

One way it might be possible to combine the acoustical advantages of the reverse fan with the practical advantages of the fan would be to segment the walls. The orientation of the segments would correspond to the reverse fan, but their positioning would provide an overall fan shape (Fig. 2.15a). As shown in Fig. 2.15b, the VSs show the same tendency as in the reverse fan to curve toward the rear, so the reverse fan and the segmented fan should have comparable SI. Ironically, segmentation of the walls *is* often applied in concert hall designs, but invariably the orientation is chosen to simulate a fan with a larger splay. The resulting central compression of the VSs is exactly the opposite of what is required to increase SI. It is intriguing to consider how the acoustics of the hall could be adjusted if these panels were allowed to rotate. By changing the degree of SI, this adjustment might provide a more meaningful control of the acoustics of a concert hall than techniques currently in vogue.

The same principles apply when considering the vertical orientation of the side walls. By reinterpreting Fig. 2.13, we see that the VSs curve upward when the walls lean in and downward when they lean out. Listeners in the orchestra seating section are at approximately the same elevation as the source, so any downward curvature will move VSs below their plane. Because the audience seating area is highly absorptive, the result will be a severe attenuation of the lateral sound, reducing SI. An upward curvature of the VSs can also be detrimental because SI decreases as sounds are elevated toward the median plane [BARRON-74]. However, a small
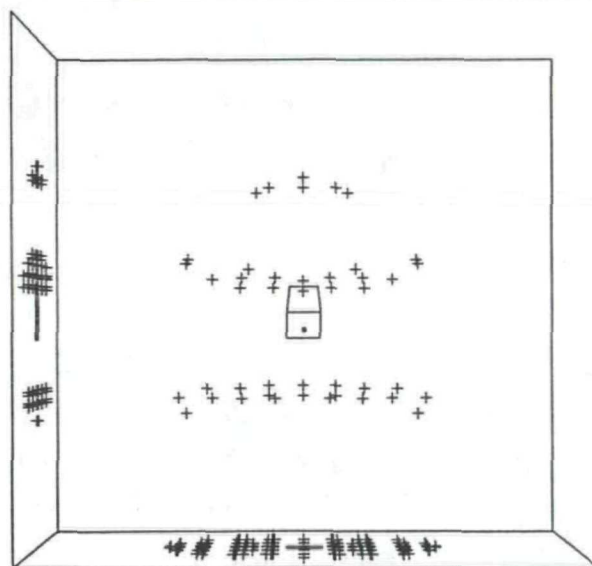
Fig. 2.16 In a "semifan" the side walls splay only part way to the rear. Even when the listener is situated near the rear of the hall the VSs show some forward curvature, as in the simple fan.

lean will not significantly reduce SI, and has the advantage of allowing the early reflections to descend on a listener, minimizing debilitating interaction with the audience. Audience members in a balcony will generally be located at a higher elevation than the source, so for them an inward lean is even more beneficial as it brings VSs up to their horizontal plane and minimizes audience absorption. One way in which this lean can be (and indeed has been) introduced uses the same concept as the segmented fan: break the walls into segments with each one positioned according to a vertical wall but with a slight downward orientation. The fronts of balconies or boxes serve admirably.

In many halls the splay of the side walls does not extend uniformly to the rear. As long as the listener is in the region of the splay, the lateral VSs will obviously correspond to a simple fan. The farther into the region of parallel walls the listener moves, the more like a rectangular hall the VSs behave. But Fig. 2.16 shows that when the splay extends halfway back, the VSs show a significant curvature even when the listener is close to the rear of the hall. The discontinuity in the side walls will also make the response of the hall less uniform.

Many studies of the acoustical implications of concert hall geometry consider only first- or second-order reflections, obviating computer modeling. The comparison in this section illustrates how important it is to consider higher-order VSs as well: detecting differences on

the basis of only the first-order reflections would have been difficult, although the differences
are apparent when higher-order VSs are considered. Thus, computer modeling is a useful tool
in analytical studies of concert hall design as well as in the creation of audible simulations.

2.2.4 Ray tracing

The image model is also useful for analyzing the characteristics of another technique often
used for computing sound paths in a concert hall, ray tracing [SCHROEDER-70][WAYMAN-
80][WALSH-81]. Ray tracing is a more popular technique for analyzing the sound paths in
a concert hall because it tends to be easier to program than the image model, particularly
for complex geometries. If it were possible for the two techniques to deal with an infinite
computation, then they would produce identical results. Differences arise as a result of the
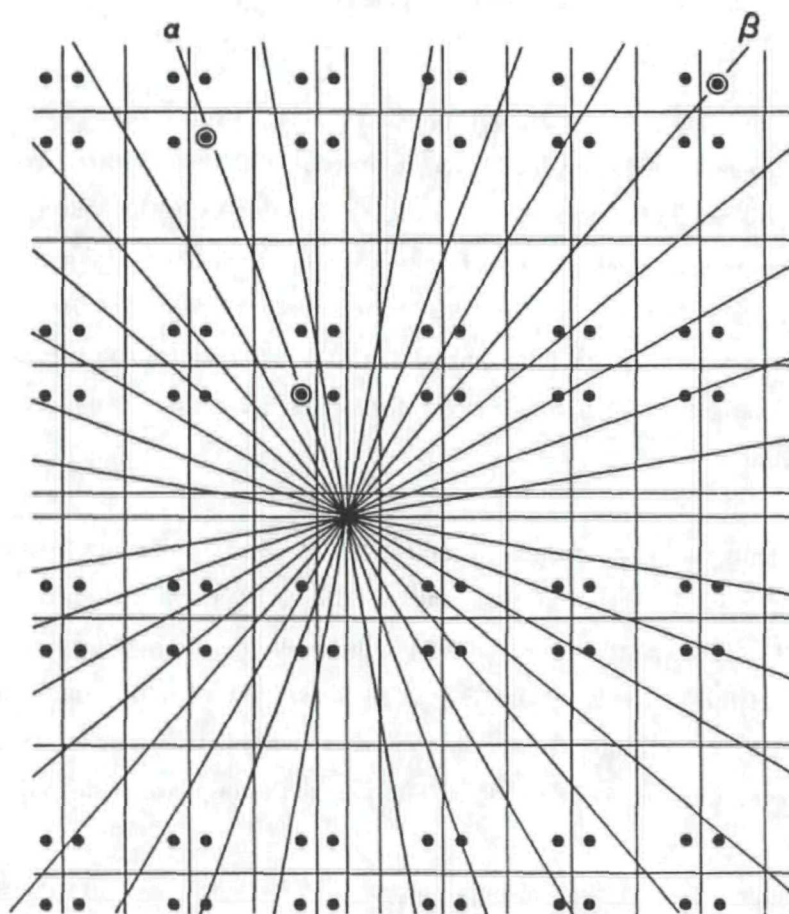
Fig. 2.17 In the ray tracing method, the rays emitted by the source are extended until they find a virtual
listener. Ray α misses the virtual listener hiding in the shadow of the closer one. Ray β finds a virtual listener
near the periphery of the display, but misses several that are closer.

different ways the methods restrict their computation to a finite size. The ray tracing method limits its computation by assuming that the source emits a finite number of rays uniformly distributed over direction. As a result, the ray tracing method will always find a certain number of sound paths, but they will not necessarily be a complete set of the sound paths within the time interval spanned. In the ray tracing method, each ray is extended by means of linear extrapolation and specular reflection until it reaches a zone of space surrounding the listener. We can analyze the behavior of the ray tracing method by using the ideas of the image model itself. Accordingly, a simple path involving a single reflection can equivalently be represented as a straight line path connecting the source to a virtual listener located in an adjacent virtual room. Once again, when we extend this idea we obtain a lattice of virtual rooms surrounding the real room (Fig. 2.17), but this lattice is the dual of the lattice in Fig. 2.1 because now each virtual room contains a virtual listener rather than a virtual source. Each of the paths that the ray tracing method finds in the real room can be unfolded into a straight line path connecting the source to one of the virtual listeners in the lattice. As usual, the image model simplifies our analysis by allowing us to consider straight line paths instead of paths involving reflections.

This figure illustrates the two principal problems that can arise in the ray tracing method. The first problem is that farther virtual listeners can be hidden behind closer ones. When the ray labeled α is emitted by the source, it travels a short distance before it finds a virtual listener located in a room only two rooms removed from its point of origin. At this point, the usual practice is to discontinue the extension of the ray. However, had it been continued it would eventually have found a second virtual listener located somewhat farther from the source. The second problem is illustrated by ray β which can be seen to slide past a number of closer virtual listeners before eventually hitting one near the periphery of the figure. Using the parameters for the ray trace shown in the figure, these closer VSs will never be found. It is important to realize that the solutions to these two problems conflict with each other. In order to solve the first we would like to reduce the size of the listener zone to minimize the size of its shadow. However, doing so will decrease the likelihood of a virtual listener's being found by a ray, exacerbating the second problem.

The image model limits its computation by rejecting virtual sources found to be too far from the listener. As a result, the image model is guaranteed to find all of the virtual sources within a certain radial distance of the listener. When the information concerning the

sound paths is reduced to the impulse response, the two methods will produce different impulse responses. All of the reflections found by the ray tracing method will correspond exactly to reflections found by the image model, but the image model will also find a number of reflections that the ray tracing method missed, and in exchange, the ray tracing method will find a number of reflections outside the period being considered. Missing some of the sound paths will lead to an underestimate of the amount of sound energy in any period of the reverberant decay. Also, the effect of omitting these early echoes will usually be clearly audible in a simulation of the concert hall. It is possible to minimize the errors of the ray tracing method by choosing a very small listener zone, emitting a very large number of rays, and then keeping only the shortest paths. The real problem is that there is no way to know whether a path has been missed. Missing even a single one can alter audible simulations. This limitation of ray tracing might also explain anomalies observed when it is used to analyze the properties of reverberation [SCHROEDER-70][WAYMAN-80].

There are many variants of the ray tracing technique described here, but they do not alter the final conclusion. For example, Wayman [WAYMAN-80] and Walsh [WALSH-81] assume the listener is a point, but the ray has cross section. It is also possible to terminate the extension of rays on the basis of their total length. Allowing rays to pass through listener zones renders them transparent. As transparent listener zones do not cast shadows, farther virtual listeners would not be lost behind closer ones. The penalty is unnecessary computation in cases where no virtual listeners were hiding. Although it might be possible to coerce the ray tracing method into providing accurate and complete information, it should be clear the the image model has the advantage of providing the desired information as a matter of course.

2.2.5 Amount of computation required

In any practical application of the extended image model it is important to have a sense of how the complexity of the model influences the amount of computation required. A convenient measure of the amount of computation is the number of VSs computed, N_{vs}. Fig. 2.18 shows empirical results for the variation in $\log_{10} N_{vs}$ vs. the truncation time t_{max} for three different concert hall geometries. In every case, the floor plan of the hall is a polygon and the cross section is rectangular. The volumes of the polyhedra were equalized to provide a consistent basis for comparison. As expected, the amount of computation increases as the number of sides
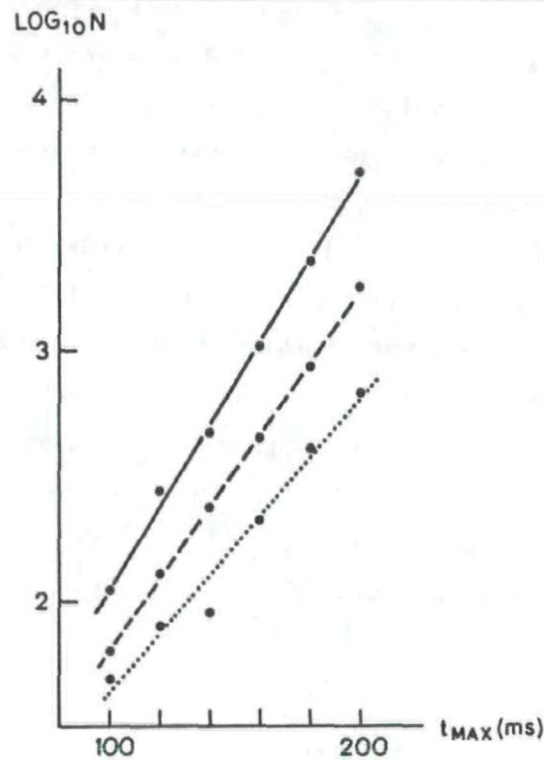
Fig. 2.18 Comparison of the amount of computation required by the extended image model for different geometries. The rooms all have the same volume and rectangular cross section, but the floor plans are polygons with different numbers of sides. (——) is for an octagonal plan, (– – –) is for a hexagonal plan, and ($\cdots$) is for a square. The approximate linearity of these traces indicates that the dependence is exponential, with a doubling in the number of VSs for an increase in t_{max} of 18, 21, and 26 ms respectively.

increases. Somewhat unexpectedly, the dependence on t_{max} is almost perfectly exponential, with a doubling of N_{vs} for an increase in t_{max} of 18–26 ms. We have found halls that produce a doubling of N_{vs} for as little as 10 ms. In such cases, one can easily initiate a computation that will require an inordinate amount of time. The adherence to an exponential law makes it possible to anticipate these situations by extrapolating from the values of N_{vs} for small t_{max} to the value for the desired t_{max}. An excessive value mandates simplification. Unfortunately, the number of sides in the polyhedron is not the only factor. Distorting the shape while maintaining the number of sides will affect the amount of computation in an unpredictable way. Even a change as simple as moving the source or listener can have a significant effect.

2.3 CONCLUSION

The extension of the image model to complex shapes eliminates its most obvious limita-

tion. The only geometrical restriction remaining is that the model of the concert hall must be piecewise planar. However, curved surfaces could be modeled using a piecewise planar approximation. The real limitation to the complexity of the shape is the computation time. By allowing the image model to deal with geometries more representative of those typically encountered in practice, it can be applied in a broader range of applications. The extended image model is the first step in creating audible simulations of concert halls, to be discussed further in subsequent chapters. It also provides insight into the fundamental acoustical properties of familiar concert hall shapes. Because the images in the fan-shaped hall curve forward toward the front of the hall, the degree of spatial impression is reduced from what is obtainable in rectangular halls. We found a geometry never applied in an actual concert hall design, the reverse fan, that would have an even greater spatial impression than rectangular halls. Other limitations of the model will be discussed in Ch. 6.

CHAPTER 3

MEASURING THE IMPULSE RESPONSE

3.0 INTRODUCTION

In measuring the impulse response of a linear system, the most direct approach is to apply an impulsive excitation to the system and observe the response. There are two basic difficulties with this approach. The first is generating the impulsive excitation, and the second is obtaining adequate dynamic range. If one is dealing with an electronic system, generating an impulse is not a severe problem. But in a concert hall it is more difficult. Although techniques exist for producing an impulsive acoustical excitation—electronic spark gaps, pistol shots, or exploding balloons are often used—assuring that the energy is equally distributed over all frequencies of interest is difficult. Little energy can be delivered to the system by an impulse: the duration must be very short to provide the desired temporal resolution, and the amplitude can be increased only to the point that the nonlinearities of the system become significant. Consequently, overcoming noise is difficult. This problem is exacerbated by a nonuniform distribution of energy in the excitation because the linearity limitation to the amplitude is usually imposed by the frequency range where most of the energy falls. In ranges that are shortchanged, it is not even possible to obtain the dynamic range that is theoretically possible.

A different approach is to excite the system with noise. Because the excitation is applied for a longer period of time, more energy is delivered to the system for a given amplitude of signal, circumventing the dynamic range problem. Further, assuring the uniformity of the energy distribution over frequency is easier. The response is the convolution of the excitation with the impulse response. The impulse response can be extracted from the measurement by crosscorrelating the noise input with the output. This chapter will discuss a particularly efficient implementation of this approach on a digital computer.

The technique to be described is similar to techniques developed in Hadamard spectroscopy by Nelson and Fredman [NELSON-70], and Harwit and Sloane [HARWIT-79]. Schroeder originally described the use of pseudorandom noise for measuring the impulse response of a concert hall in [SCHROEDER-79b]. In this work, he performed the crosscorrelation operation in the frequency domain. Because the length of the pseudorandom noise sequence was one less than a power of 2, it was necessary to interpolate a sample in order to exploit the FFT. Cabot also reported using a crosscorrelation technique based on analog bucket-brigade delay lines [CABOT-79]. Basing the measurement technique on a digital computer offers the greatest potential for maximizing the dynamic range. Also, having the final result in digital form makes additional manipulations more convenient.

Because the techniques are not widely known, this chapter will begin with a brief review of the relevant properties of maximal-length sequences which are used as the pseudo noise excitation. After reviewing the crosscorrelation operation, we will show how the problem can be manipulated into a form that makes it possible to apply an efficient algorithm known as the fast Hadamard transform. Finally, we will illustrate the application of the technique, and discuss some practical issues.

3.1 MAXIMAL LENGTH SEQUENCE

The first step of the crosscorrelation method is to generate a noise-like test signal to be applied to the linear system. Although several techniques exist for generating noise in a computer [NELSON-70][HARWIT-79][KNUTH-81][DAVIES-66][RADER-70] [MACWILLIAMS-76], binary maximal-length shift register sequences have a number of advantages. The most
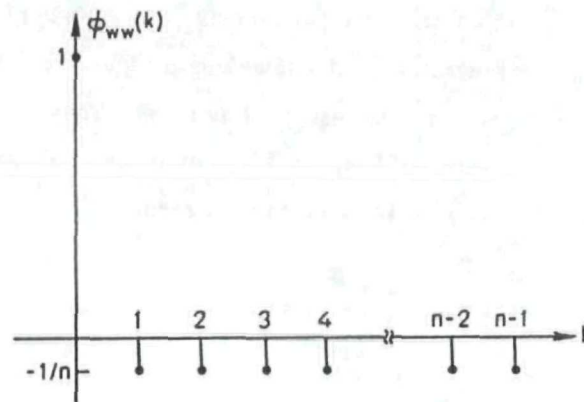
Fig. 3.1   Autocorrelation function of a maximal-length sequence with length n.

important is that except for a small DC error, their autocorrelation is a perfect impulse (see Fig. 3.1)[SARWATE-80][MACWILLIAMS-76]. In other words, the spectrum of the pseudo noise is flat everywhere except at zero frequency. Although a maximal-length sequence is actually a deterministic signal, it preserves this desirable property of white noise. Another advantage of maximal-length sequences is that they are very easy to generate, requiring a minimum of execution time on a general-purpose computer, or of hardware in a specialized device. Also, because the sequences are binary, the crosscorrelation operation is particularly simple. By assigning the values ± 1 to the two binary levels, it is clear that the crosscorrelation requires no multiplications, only additions and subtractions. (For simplicity, no distinction will be made in what follows between additions and subtractions in discussing computational requirements because the requirements as measured in execution time or in hardware are so similar.) Multiplication usually requires much more time than addition, so eliminating the need for multiplications greatly speeds the processing. But even more significantly, an efficient algorithm based on the fast Hadamard transform (FHT) exists for performing the additions. Like the more familiar FFT, the FHT requires on the order of $n \log_2 n$ operations, which in this case are additions. So basing the technique on excitation by a maximal-length sequence makes it possible to perform the crosscorrelation expeditiously. A final advantage is that, although maximal-length sequences have statistical properties like true white noise, they are actually deterministic signals that can be repeated precisely. As long as the system is time-invariant, differences in its response can be unambiguously attributed to noise in the system.

Several of the references explain how to generate maximal-length sequences and provide a

mathematical framework based on primitive polynomials [NELSON-70][HARWIT-79][DAVIES-66][MACWILLIAMS-76]. The generation of maximal-length sequences is most easily described by considering a specific case, such as the 3-stage shift register shown in Fig. 3.2. The boxes containing z^{-1} represent a unit-sample delay produced by memory elements or flip-flops. The operation designated by the $\oplus$ is a modulo-two sum, or exclusive-or, defined by

$$
\begin{aligned}
0 \oplus 0 &= 0 \\
0 \oplus 1 &= 1 \\
1 \oplus 0 &= 1 \\
1 \oplus 1 &= 0.
\end{aligned}
\tag{3.1}
$$

A signal is fed back to the beginning of the shift register which is a modulo-two sum of selected outputs. In other words, the shift register generates a sequence of ones and zeros that satisfies the recursion relation

$$
\tilde{w}(k+3) = \tilde{w}(k) \oplus \tilde{w}(k+2).
\tag{3.2}
$$

Recursion relations can be specified for any shift register length. The sequences produced at each node of the 3-stage shift register in Fig. 3.2 that are shown in the figure were produced by initializing the shift register to all ones.   Choosing different initial conditions will change the sequences produced in a way that corresponds to delaying the sequences by some amount. With m stages in the binary shift register it is theoretically possible to describe 2^m states, but if the content of the shift register is all zeros, it will be impossible for a one to occur, and the shift register will remain frozen in this state. In order to avoid this degenerate case, the longest sequence that can be generated using linear feedback has length $2^m - 1$. A binary sequence whose length is $2^m - 1$ is called a maximal-length sequence, or m-sequence.
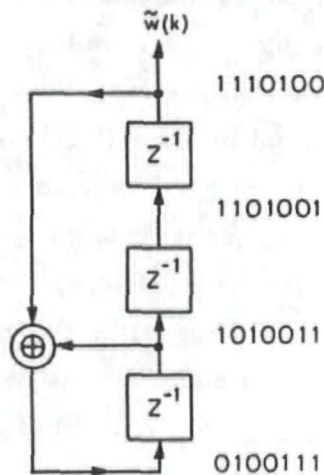
Fig. 3.2   Binary feedback shift register of length $m = 3$ for generating a maximal-length sequence of length $n = 7$.

3.2   CROSSCORRELATION

In order to drive the linear system under test, the Boolean sequence produced by the shift register must be converted to a signal. It is customary in describing digital circuits to denote the two logical levels as 1 and 0. For signal processing, the sequence $\tilde{w}(k)$ produced by the shift register is mapped to $w(k)$ by changing 1 to -1 and 0 to $+1$. (Any vector or matrix augmented with the tilde refers to the 1,0 code.) The sequence $w(k)$ is the one actually applied to the system. The crosscorrelation of the input and the output is related to the autocorrelation of the input by a convolution with the impulse response:

$$\phi_{wy}(k) = \phi_{ww}(k) * h(k). \tag{3.3}$$

When the input autocorrelation $\phi_{ww}(j) = \delta(j)$, the Dirac delta function, the result of convolving $\phi_{ww}(j)$ with any function is the function itself, in this case, the impulse response. Thus, the impulse response can be recovered by crosscorrelating the noise input $w(k)$ with the output $y(k)$ [OPPENHEIM-75]. The desirable impulsive autocorrelation of m-sequences arises only under circular autocorrelation, so the indexing of the sequences in any correlation operation must be performed modulo n. Using the notation $((j))_n$ for the residue of j modulo n, or simply $((j))$ where the modulus can be inferred from the context, the crosscorrelation operation is defined by

$$\phi_{wy}(k) = \frac{1}{n} \sum_{j=0}^{n-1} w(j)y((j+k)). \tag{3.4}$$

By a change of indices, this expression is equivalent to

$$\phi_{wy}(k) = \frac{1}{n} \sum_{j=0}^{n-1} w((j-k))y(j). \tag{3.5}$$

The circularity of the operation can also be achieved by performing linear crosscorrelations with periodic versions of the original sequences defined by

$$x_p(k) = x((k))_n. \tag{3.6}$$

In other words, each period of the periodic sequence is equal to the original sequence. Eq. (3.5) can also be described in terms of a matrix multiplication:

$$\Phi_{wy} = \frac{1}{n}\mathbf{W}_n Y. \tag{3.7}$$

Φ_{wy} and Y are vectors whose elements are the $\phi_{wy}(\cdot)$ and $y(\cdot)$ of Eq. (3.5), and the matrix $\mathbf{W}_n$ contains the circularly delayed versions of the sequence $w(\cdot)$. The noise matrix $\mathbf{W}_n$ is a right circulant matrix [GRAY-77] because successive rows are obtained from the previous one by rotating it one position to the right. For the specific case of $m = 3$, the noise matrix is

$$\mathbf{W}_7 = \begin{bmatrix} - & - & - & + & - & + & + \\ + & - & - & - & + & - & + \\ + & + & - & - & - & + & - \\ - & + & + & - & - & - & + \\ + & - & + & + & - & - & - \\ - & + & - & + & + & - & - \\ - & - & + & - & + & + & - \end{bmatrix} \qquad (8)$$

where the shorthand "+" and "−" has been substituted for $+1$ and -1.

Because the elements of $\mathbf{W}_n$ are all ± 1, only additions and subtractions are required to perform the matrix multiplication. Finding each element of the correlation vector Φ requires $n-1$ additions. There are n elements in the result vector, so the total number of additions required is $n(n-1)$. When n is a large number, the number of operations can become prohibitively large. In architectural acoustics one is interested in measuring the impulse response of concert halls. The duration of the impulse response, as defined by the reverberation time, is often as long as 2–3 seconds. In order to deal with the entire audio bandwidth of 20 Hz–20 kHz, a sampling rate of at least 40 kHz is required. As a result, the number of samples in the m-sequence could be on the order of 10^6, in which case the total number of operations would be on the order of 10^{10}. Assuming a computer is capable of performing an addition in $1\mu s$, several hours of computer time would be required. Clearly, a more efficient algorithm must be found when one is interested in dealing with such long sequences.

3.3   FAST HADAMARD TRANSFORM

The efficient algorithm for performing the desired crosscorrelation is based on the fast Hadamard transform. Like the discrete Fourier transform, the Hadamard transform can be described in terms of a matrix multiplication. The matrix that transforms the input vector

is known as the Hadamard matrix, $\mathbf{H}_n$, where n gives the number of rows or columns. The elements of the Hadamard matrix are all ± 1, and the matrix must satisfy the relation

$$\mathbf{H}_n \mathbf{H}_n^T = n \mathbf{I}_n. \tag{3.9}$$

The efficient algorithm applies only to the specific class of Hadamard matrices known as Sylvester type. The Sylvester-type Hadamard matrix is defined recursively by

$$\mathbf{H}_1 = [1]$$
$$\mathbf{H}_{2i} = \begin{bmatrix} \mathbf{H}_i & \mathbf{H}_i \\ \mathbf{H}_i & -\mathbf{H}_i \end{bmatrix}. \tag{3.10}$$

Only orders 2^k, where k is a nonnegative integer, exist. The Sylvester-type Hadamard matrix of order 8 is

$$\mathbf{H}_8 = \begin{bmatrix} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ + & + & + & + & - & - & - & - \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \\ + & - & - & + & - & + & + & - \end{bmatrix}. \tag{3.11}$$

The partitioning emphasizes the structure that evolves from the recursive definition.

The flow graph for an 8-point FHT is shown in Fig. 3.3. It should be evident that the flow graph is identical to the flow diagram for the FFT except that the twiddle factors [RABINER-74] are all unity, reflecting the fact that no multiplications are required. It should also be noted that the bit-reversal shuffling of the input vector for the decimation in time algorithm or the output sequence for the decimation in frequency algorithm is not required in the FHT. When the bit reversal is performed, the transformation is often called the fast Walsh transform (FWT) because the transform terms fall in order of increasing sequency [HARMUTH-69].

A $\longrightarrow$ A + B

B $\longrightarrow$ A - B

BASIC BUTTERFLY ELEMENT

a $a+b+c+d+e+f+g+h$

b $a-b+c-d+e-f+g-h$

c $a+b-c-d+e+f-g-h$

d $a-b-c+d+e-f-g+h$

e $a+b+c+d-e-f-g-h$

f $a-b+c-d-e+f-g+h$

g $a+b-c-d-e-f+g+h$

h $a-b-c+d-e+f+g-h$

a	$a+b$	$(a+b)+(c+d)$	$[(a+b)+(c+d)]+[(e+f)+(g+h)]$
b	$a-b$	$(a-b)+(c-d)$	$[(a-b)+(c-d)]+[(e-f)+(g-h)]$
c	$c+d$	$(a+b)-(c+d)$	$[(a+b)-(c+d)]+[(e+f)-(g+h)]$
d	$c-d$	$(a-b)-(c-d)$	$[(a-b)-(c-d)]+[(e-f)-(g-h)]$
e	$e+f$	$(e+f)+(g+h)$	$[(a+b)+(c+d)]-[(e+f)+(g+h)]$
f	$e-f$	$(e-f)+(g-h)$	$[(a-b)+(c-d)]-[(e-f)+(g-h)]$
g	$g+h$	$(e+f)-(g+h)$	$[(a+b)-(c+d)]-[(e+f)-(g+h)]$
h	$g-h$	$(e-f)-(g-h)$	$[(a-b)-(c-d)]-[(e-f)-(g-h)]$
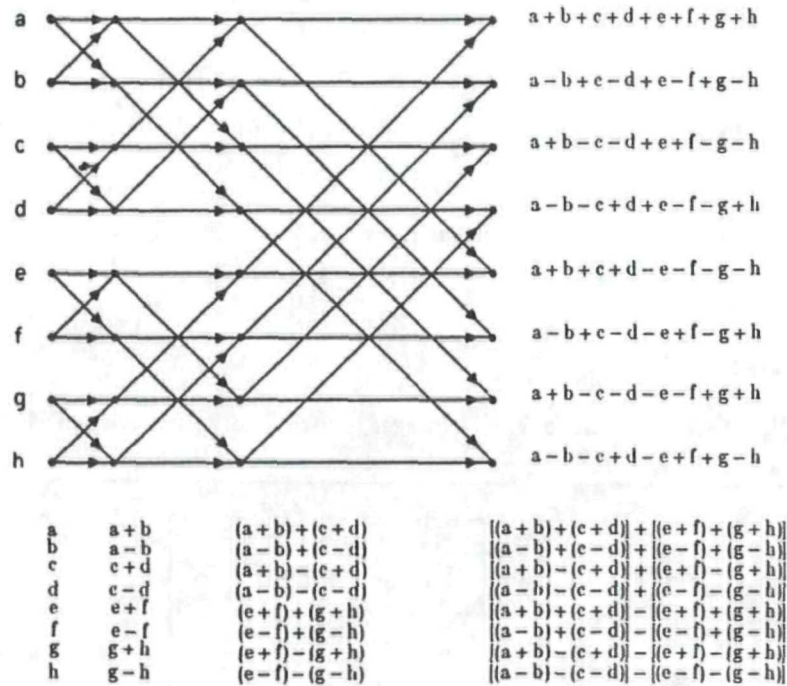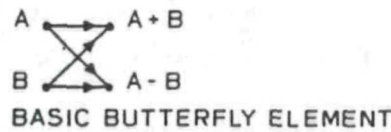
Fig. 3.3   Flow graph for an 8-point fast Hadamard transform.

3.4   PERMUTATIONS

3.4.1   Using the FHT to perform the crosscorrelation

The crosscorrelation operation is described in terms of a matrix multiplication in Eq. (3.7). Unfortunately, $\mathbf{W}_n$ is not a Hadamard matrix, so it is not possible to apply the FHT directly. But it is possible to manipulate the problem into the required form by a sequence of matrix multiplications:

$$\mathbf{W}_n = \mathbf{P}_2 \mathbf{S}_2 \mathbf{H}_{n+1} \mathbf{S}_1 \mathbf{P}_1 \tag{3.12}$$

$\mathbf{P}_1$ and $\mathbf{P}_2$ are permutation matrices whose purpose is to permute the rows and columns of $\mathbf{H}_{n+1}$. Matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ suppress the first row and column of $\mathbf{H}_{n+1}$ in order to reduce the $(n+1) \times (n+1)$ matrix into one that is only $n \times n$. The reader may confirm this equivalence

for the specific case of $n = 7$ by using the matrices

$$
\mathbf{P}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{3.13}
$$

$$
\mathbf{S}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{S}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \tag{3.14}
$$

By substituting Eq. (3.12) into Eq. (3.7), we obtain the relation

$$
\Phi_{wy} = \frac{1}{n} \mathbf{P}_2 (\mathbf{S}_2 (\mathbf{H}_{n+1} (\mathbf{S}_1 (\mathbf{P}_1 Y)))). \tag{3.15}
$$

Although the operations can be performed in any order, the sequence indicated by the parentheses leads to the following interpretation. The measurement vector Y is permuted according to $\mathbf{P}_1$ and a 0 element is affixed to the beginning of the vector. The resulting vector, which has the required 2^m terms, is transformed by the FHT algorithm. Then the first element is dropped, and the final result is obtained by permuting the vector according to $\mathbf{P}_2$ and normalizing. This sequence of operations is illustrated for the specific case of a sequence with length 7 in Fig. 3.4.

3.4.2 Determining the permutation matrices

Once one knows the permutation matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, their validity can always be checked by assuring that Eq. (3.12) holds. To find the permutations in the first place, we start with the $\mathbf{W}$ matrix and work backwards to the $\mathbf{H}$ matrix. The permutations to be described here are

$$Y = \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \end{bmatrix} \quad P_1 Y = \begin{bmatrix} a \\ f \\ b \\ g \\ e \\ d \\ c \end{bmatrix} \quad S_1 P_1 Y = \begin{bmatrix} 0 \\ a \\ f \\ b \\ g \\ e \\ d \\ c \end{bmatrix}$$

$$H_8 S_1 P_1 Y = \begin{bmatrix} 0 + a + f + b + g + e + d + c \\ 0 - a + f - b + g - e + d - c \\ 0 + a - f - b + g + e - d - c \\ 0 - a - f + b + g - e - d + c \\ 0 + a + f + b - g - e - d - c \\ 0 - a + f - b - g + e - d + c \\ 0 + a - f - b - g - e + d + c \\ 0 - a - f + b - g + e + d - c \end{bmatrix}$$

$$P_2 S_2 H_8 S_1 P_1 Y = \begin{bmatrix} -a - b - c + d - e + f + g \\ +a - b - c - d + e - f + g \\ +a + b - c - d - e + f - g \\ -a + b + c - d - e - f + g \\ +a - b + c + d - e - f - g \\ -a + b - c + d + e - f - g \\ -a - b + c - d + e + f - g \end{bmatrix} = W_7 Y$$

Fig. 3.4   Illustration of the efficient algorithm for the specific case of a sequence of length $n = 7$.

not essentially different from those first presented by Nelson and Fredman [NELSON-70]. The relation we seek to establish is derived from Eq. (3.12):

$$G_n = S_2 H_{n+1} S_1 = P_2^{-1} W_n P_1^{-1}. \tag{3.16}$$

The new matrix G_n is the original H_{n+1} matrix with its first row and column dropped. The procedure to be described will actually produce the inverse of the permutation matrices, but inverting a permutation matrix is accomplished simply by transposing it. The derivation of the permutations is facilitated by dealing with $\tilde{W}$ as well as W, and seeking permutation matrices that transform it to $\tilde{G}$. A parallel sequence of operations applied to W will transform it to G. P_1^{-1} describes a permutation of the columns of the W matrix. The columns of $\tilde{W}$ are versions of the same maximal-length sequence obtained by progressively delaying each column. The different versions can also be considered to arise from different initial conditions in the shift register that generates the sequences. Once the initial conditions have been specified, all of the $2^m - 1 = n$ terms of the sequence are predestined. Therefore, the identity of each column can be established by using any m consecutive terms of the column. By assigning binary weights to the terms, a number is produced that distinguishes the columns, and so serves as a tag. In computing the tag, the significance of the terms can be defined in either order. The decision about which m consecutive terms to consider and which order of significance to assign is based

$$
\tilde{W} = \begin{array}{ccccccc} 1 & 3 & 7 & 6 & 5 & 2 & 4 \end{array} \quad
\left[\begin{array}{ccccccc}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 1 \\
\hdashline
1 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 & 1
\end{array}\right]
\;\rightarrow\;
\tilde{W}' = \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \quad
\left[\begin{array}{ccccccc}
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\hdashline
0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 & 0
\end{array}\right]
$$

Fig. 3.5 Permutation matrix P_1^{-1} rearranges the columns of $\tilde{W}$ so that the tags are in ascending numerical order.

on programming convenience. The matrix P_1^{-1} is defined to be the matrix that permutes the columns of W so that the keys are in ascending numerical order. Performing this operation produces the matrix $\tilde{W}'=\tilde{W}P_1^{-1}$. The operations performed so far are illustrated in Fig. 3.5.

Now we must determine the second permutation. The matrix P_2^{-1} is the matrix that permutes the rows of $\tilde{W}'$ to $\tilde{G}$. In order to derive this permutation, we start by examining H for a way to tag each row in a manner analogous to the procedure used before for rearranging the columns. The nature of the tag is suggested by the recursive definition for the Sylvester-type Hadamard matrix, Eq. (3.10). This definition indicates that the first entry in each row of H_{2i} is simply the Hadamard matrix of the next lower order, H_i. The second entry is also obtained from H_i either by direct replication, or by inversion. We can use a single bit to record which operation was performed, with 0 indicating direct replication and 1 indicating inversion. Thus, the first row of each H_{2i} has a 0 associated with it, and the second row has a 1. By placing the bit in the ith position of a binary word, a word is generated that tags each row. This procedure is illustrated in Fig. 3.6.

Fig. 3.6 Illustration of how the recursive definition of the Sylvester-type Hadamard matrix leads to the development of a tag for each row.

$$
\tilde{\mathbf{W}}' =
\begin{bmatrix}
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 & 0
\end{bmatrix}
\begin{matrix}
1 \\ 2 \\ 4 \\ 3 \\ 6 \\ 7 \\ 5
\end{matrix}
\rightarrow
\tilde{\mathbf{G}} =
\begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}
\begin{matrix}
1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7
\end{matrix}
$$

Fig. 3.7   Permutation matrix $\mathbf{P}_2^{-1}$ rearranges the rows of $\tilde{\mathbf{W}}'$ so that the tags are in ascending numerical order.

To determine $\mathbf{P}_2^{-1}$, we examine the entries of $\tilde{\mathbf{W}}'$ to deduce the corresponding tag. The desired permutation matrix is defined to be the matrix that arranges the tags in ascending numerical order. Examining $\tilde{\mathbf{W}}'$ in Fig. 3.7 shows that the tag can be assembled by collecting the m terms from each row at the positions 2^k, where $k = 0, 1, \ldots, (m-1)$. Performing this operation will produce the matrix $\tilde{\mathbf{G}}_n$, as expected (Fig. 3.7). Applying the same permutations to $\mathbf{W}_n$ will result in $\mathbf{G}_n$.

In a practical implementation, the permutations can be generated once in a separate operation from the crosscorrelation and stored on disk. By this means, repeated crosscorrelations for the same sequence length do not require that the permutations be generated repeatedly.

3.5   DEMONSTRATION

The algorithm can be demonstrated by measuring the impulse responses of known systems. Fig. 3.8 shows the measured impulse responses for a simple delay, and an 8th-order Butterworth low-pass filter. These measurements were obtained using a maximal-length sequence of length $n = 1023$. Aside from a small DC error (see Sec. 3.6.2) that is imperceptible on the scale of the plots, the impulse responses are exactly correct. One should be aware that if the measurements had been performed using random noise, there would have been some uncertainty in the amplitude of the results obtained. The actual impulse responses would only be approached asymptotically as the measurement time increased without bound. The near perfection using pseudorandom noise arises from the fact that the excitation is deterministic. It should perhaps be emphasized that in a practical application there will still be uncertainty in the amplitude
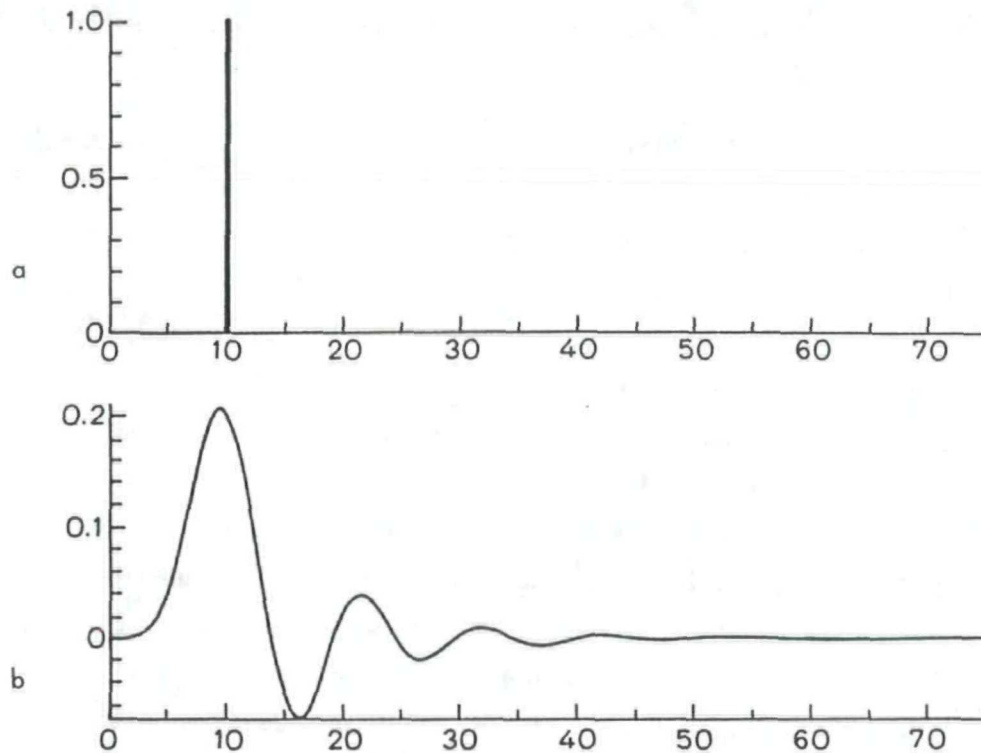
Fig. 3.8 Computed impulse response of (a) a simple delay of 10 samples, and (b) an 8th-order Butterworth low-pass filter.

due to noise in the system under test. But unlike measurements based on truly random noise, the excitation itself does not contribute to this randomness.

Because the measurement is discrete time, the temporal resolution is limited by the sample period. Consequently, there is always a small uncertainty in the group delay. In some cases there is also uncertainty in the amplitude because the impulse response is only determined to the number of points in the excitation. However, as long as the system is strictly bandlimited to the Nyquist frequency, the intermediate values are available by interpolation.

3.6 PRACTICAL CONSIDERATIONS

3.6.1 Choosing the length of the m-sequence

When the m-sequence is injected into a linear system, the output of the system is the con-

volution of the input with the impulse response. The convolution is described mathematically by

$$y(k) = w(k) * h(k) = \sum_{j=0}^{n-1} w((k-j))h(j). \tag{3.17}$$

As mentioned previously, the desirable autocorrelation property of m-sequences arises only under circular operations. But as a practical matter, dealing with a linear convolution is usually more straightforward. Fig. 3.9 gives a pictorial representation of the convolution operation. The periodic pseudo noise excitation is delayed and then weighted by the corresponding term of the impulse response. The final output is obtained by summing all of the partial results. It can be seen in this figure that the effect of circular convolution can be obtained by retaining the data from any period of the output after the first. This illustration also clearly shows the effect of attempting to deconvolve an impulse response that is longer than the input m-sequence. If there were an additional term in the impulse response, $h(7)$, its delayed sequence would overlap the repetition of the first undelayed term. There is no way of resolving the ambiguity, so an error will result. The error will be insignificant, though, if the amplitude of the impulse response has decayed to a small value. Clearly no error results if the length of the m-sequence is longer than the impulse response, but there is a disadvantage in the additional computation required. For a finite impulse response (FIR) filter, the overlap error can always be avoided by choosing the duration of the pseudo noise excitation to exceed that of the impulse response. But for an infinite impulse response (IIR) filter, there will always be some overlap. The severity of the
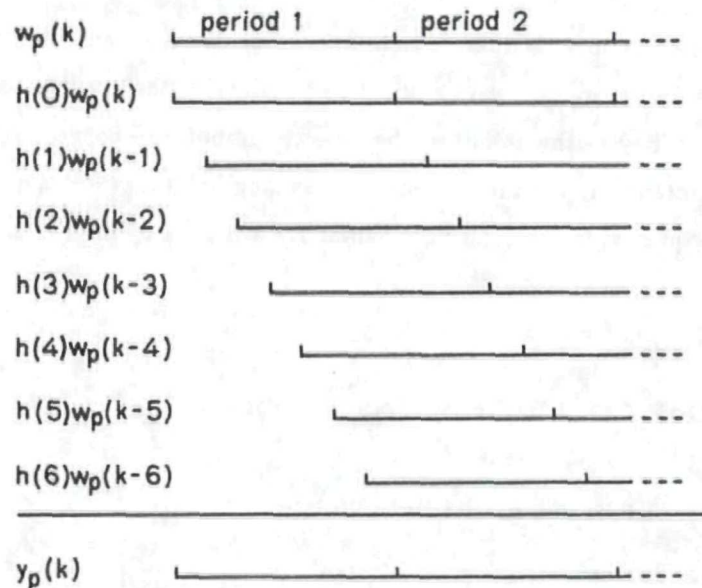
Fig. 3.9   Pictorial representation of the convolution operation showing the relative timing of the partial results.

error that results can be reduced by choosing the duration of the pseudo noise sequence to be long enough that the amplitude of the impulse response has decayed to a small value.

3.6.2   Effect of DC error in autocorrelation

As shown in Eq. (3.3), the convolution also relates the input autocorrelation to the crosscorrelation. This relationship is what makes it possible to recover the impulse response by way of the crosscorrelation when the input autocorrelation is a delta function. Unfortunately, the autocorrelation of the m-sequence is not a perfect impulse, as can be seen in Fig. 3.1. Rather, the autocorrelation is actually

$$\phi_{ww}(k) = \frac{n+1}{n}\delta(k) - \frac{1}{n}. \tag{3.18}$$

Thus, the crosscorrelation has an error that can be found as follows:

$$
\begin{aligned}
\phi_{wy}(k) &= \sum_{j=0}^{n-1} h(j)\left[\frac{n+1}{n}\delta((j-k)) - \frac{1}{n}\right] \\
&= \frac{n+1}{n}\sum_{j=0}^{n-1} h(j)\delta((j-k)) - \frac{1}{n}\sum_{j=0}^{n-1} h(j) \\
&= \frac{n+1}{n}h(k) - \frac{1}{n}\sum_{j=0}^{n-1} h(j).
\end{aligned}
\tag{3.19}
$$

The summation term in Eq. (3.19), which is the DC component of the actual impulse response, is also the DC error in the computed impulse response. We can calculate the error in the computed impulse response by adding together the terms of the crosscorrelation:

$$
\begin{aligned}
\sum_{k=0}^{n-1}\phi_{wy}(k) &= \frac{n+1}{n}\sum_{k=0}^{n-1} h(k) - \sum_{j=0}^{n-1} h(j) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} h(k).
\end{aligned}
\tag{3.20}
$$

The impulse response can be found exactly by correcting the crosscorrelation:

$$h(k) = \frac{n}{n+1}\left[\phi_{wy}(k) + \sum_{j=0}^{n-1}\phi_{wy}(j)\right]. \tag{21}$$

In practice, this correction is rarely necessary. Most practical systems in audio have no transmittance at DC, so that the error term in Eq. (3.19) is nil. As long as the impulse response does not have a strong DC component, the error term will be negligible.

3.6.3 Additional improvement of dynamic range

When measuring the impulse response of a linear system using an impulsive excitation, a technique often used to improve the signal-to-noise ratio is to average together the response to a number of impulses [BERMAN-77]. The desired response adds coherently, but the noise adds incoherently, providing an improvement in SNR of $\sqrt{N}$, where N is the number of periods being averaged together. This procedure has the disadvantage of consuming considerable time because one must wait until the system has returned to its quiescent state before initiating the next measurement. If the impulse response has a duration of T seconds, a total of NT seconds will be required for the measurement. The crosscorrelation method starts with a tremendous advantage in dynamic range. The advantage is exactly the same as what would be obtained by averaging together n impulse responses, but the measurement only requires time T to perform. However, in extremely noisy environments, it would be possible to combine the averaging technique with the crosscorrelation method in order to obtain an additional improvement in dynamic range. Schroeder reports [SCHROEDER-79b] that this technique has been used to measure the impulse response of a lecture hall while the hall was being used by exciting the hall with a quiet pseudo noise signal. It might be possible to apply the same concept to measuring the response of a concert hall during a performance while an audience is present. This approach would circumvent the difficult problem of relating the behavior of the empty hall to its response with an audience present.

Averaging many periods trades smaller storage requirements for longer measurement times. The pseudo noise must have a duration at least as great as the duration of the impulse response to avoid overlap distortion (Sec. 3.6.1). Longer sequences improve the dynamic range, with each doubling of the sequence length doubling the dynamic range in logarithmic terms, i.e., doubling the number of dB. Averaging two periods also doubles the number of samples of excitation, but improves the dynamic range by only 3 dB. However, averaging improves the dynamic range without increasing the storage requirements. As long as storage is available one clearly should opt for a longer m-sequence as the rate of increase in dynamic range is much greater.

In addition to consuming less time, the crosscorrelation method offers another important advantage over averaging in many practical situations. For an averaging technique to work, the desired signal must add coherently every period. But in a concert hall, the impulse response slowly changes due to small changes in the speed of sound, so measurements extending over long periods will fail unless special steps are taken. The speed of sound is a function of temperature and humidity, and is also affected by air currents. Beranek [BERANEK-54] gives an expression for the speed of sound c as a function of temperature θ:

$$c(\theta) = 331.4 + 0.607\theta \text{ m/s}. \tag{3.22}$$

Suppose the measurement is made with a sampling rate of 44.1 kHz and that an accuracy of half a sample period is desired at 100 ms. The sound travels the same distance regardless of its speed, so we have

$$c(\theta)\tau = c(\theta')\tau'. \tag{3.23}$$

Using these two equations, one can show that a decrease in τ of half a sample period corresponds to an increase in temperature of only 0.14° C. As we are often interested in measuring the impulse response at much longer times—several seconds perhaps—the temperature excursions that normally occur in an occupied hall greatly exceed what can be tolerated. Therefore, averaging would usually require estimating the time-base variation and resampling the data to make the desired signals coincide. The crosscorrelation method usually circumvents this problem because its time frame is short compared to variations in the response of the hall.

One situation in which the crosscorrelation method *can* have trouble with time-base variations is when the response is tape recorded. Tape recording minimizes the amount of equipment that must be transported to remote locations. The pseudo noise is easily generated with a simple circuit or tape recorded on a second machine. The recorded response is returned to the laboratory and fed into the computer for processing. Complications arise due to the imprecision in the speed of tape recorders. Analog tape recorders have wow and flutter. Digital tape recorders have essentially no wow and flutter, but their sampling rates are never exact. As a result, the excitation will be produced at one sampling rate and the response recorded at a slightly different one. Naturally, a technique based on crosscorrelation will fail when the two sequences correspond to different sampling rates. (Ironically, an impulsive excitation produces a measurement much more robust to this distortion.) To illustrate the seriousness of the problem,

two tape recorders whose sampling rates differ by only 50 ppm will be nearly a sample apart after one period when the shift register length $m = 14$. As m might have to be as large as 17 to measure a concert hall with full bandwidth, the problem can be serious indeed. Our solution was to drive a pseudo noise generator with the clock of the digital tape recorder, assuring that the two were perfectly synchronized. By this means we were able to obtain measurements that were very accurate and repeatable (Fig. 3.10).

3.6.4   Estimate of computational requirements

It is more difficult to develop a convenient rule of thumb for the number of computations required by the FHT than for the FFT. In analyzing the FFT it is customary to consider only the number of multiplications, because they usually dominate the computations. However, the elemental operation in the FHT is addition, which will not dominate the other computations required. Each stage of the FHT requires n additions for the butterfly operation. Since there are $\log_2 n$ stages, the total number of additions required by the butterfly to evaluate an n-point FHT is $n \log_2 n$. However, in addition to the butterfly, it is also necessary to compute the indices of the two terms being applied to the butterfly—two more additions—and to test for the completion of a loop. The compare operation in the loop completion test is basically a subtraction, so 5 additions are required for every two points, increasing the number of additions
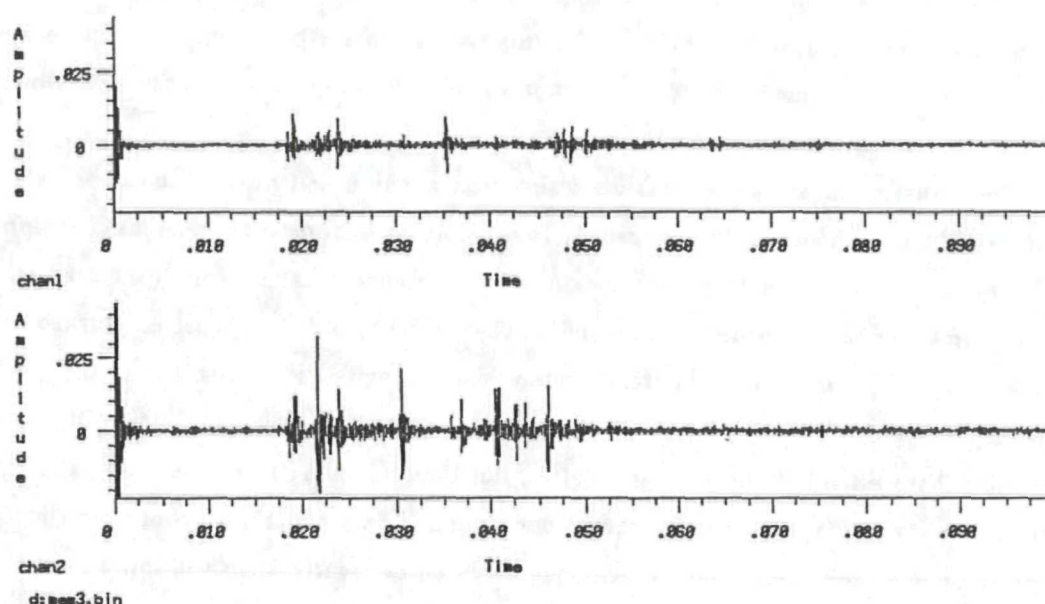
Fig. 3.10 A portion of the binaural impulse response of a concert hall as measured with the crosscorrelation technique. Several of the early reflections are clearly visible.

per stage from n to $2.5n$. Disregarding assignments, the best estimate for the number of operations required is $2.5n \log_2 n$. Note that in the example we considered previously of a 3-second sequence sampled at 40 kHz, the number of additions is reduced from approximately 10^{10} to about 5×10^6, a decrease of more than 10^3! Assuming the same computation rate as before of 10^6 additions per second, the time required to compute the impulse response will be reduced to the much more manageable figure of about 6 sec.

3.6.5 Calculating the frequency response

One way to obtain the frequency response of the system under test is to apply the method described in this chapter to determine the impulse response, and then to transform the impulse response to the frequency domain using a DFT. But if one is interested only in the frequency response, then another approach worth considering is to perform the computations in the frequency domain. It is well known that the correlation of two sequences can be related to a convolution:

$$\phi_{wy}(k) = w(-k) * y(k). \tag{3.24}$$

The z-transform of the convolution of two sequences is the product of their z-transforms

$$w(k) * y(k) \leftrightarrow W(z)Y(z). \tag{3.25}$$

We also have the relation

$$x(-k) \leftrightarrow X(\frac{1}{z}). \tag{3.26}$$

Combining Eqs. (3.24)–(3.26) we obtain

$$\phi_{wy}(k) \leftrightarrow W(\frac{1}{z})Y(z). \tag{3.27}$$

Because we are interested in the frequency response, we substitute $z = e^{j\omega}$ to obtain finally

$$\phi_{wy}(k) \leftrightarrow W(e^{-j\omega})Y(e^{j\omega}) = W^*(e^{j\omega})Y(e^{j\omega}). \tag{3.28}$$

In practice, the Fourier transform would be computed as a DFT. All of the terms of $W^*(e^{j\omega})$ can be calculated beforehand and stored. To find the frequency response one need only transform

the measurement $y(k)$, and multiply its transform term-by-term with the stored values of W^*. Note that although the magnitude of $W^*(e^{j\omega})$ will be constant (except for the DC term), the phase will vary, so the multiplications will not be trivial.

As far as the amount of computation is concerned, the transformation of $y(k)$ required by the frequency domain approach is equivalent to the transformation of the impulse response in the time domain approach. In addition, the frequency domain approach requires n complex multiplications ($4n$ real multiplications), and the time domain approach requires $2.5n \log_2 n$ real adds. To compare the amount of time required for the two approaches, we can suppose that multiplications take 10 times as long as adds, and find the value of n at which the computation times are equivalent:

$$2.5n \log_2 n = 10 \times 4 \times n$$
$$\log_2 n = 16 \quad \rightarrow \quad n = 2^{16} \tag{3.29}$$

For $n > 2^{16}$ the frequency domain approach will require less time, and for $n < 2^{16}$ the time domain approach is faster. If one is interested in the impulse response, it would also be possible to find the frequency response using the frequency domain approach, and inverse Fourier transforming the result. Clearly, this approach will require a great deal more computation than the algorithm presented in this chapter that finds the impulse response directly in the time domain.

3.7 SUGGESTIONS FOR ADDITIONAL WORK

The length of the impulse response that can be recovered by crosscorrelation is limited by the amount of available main memory. Two arrays must be present simultaneously, each as large as the m-sequence. For acoustical measurements it is often of interest to measure impulse responses as long as 2–3 seconds. In order to deal with the entire audio spectrum up to 20 kHz, the sampling rate must be at least 40 kHz. Accordingly, each array could be as long as 120 000 samples, requiring that 240 000 samples be stored in main memory. In computer systems which do not have such prodigious memory capacity, or in cases when one is interested in measuring longer impulse responses, or when the sampling rate must be considerably higher, the resulting abundance of samples might overwhelm the resources available. In such cases it would be desirable to modify the algorithm to reduce the memory requirements.

There is a fairly straightforward way to halve the memory requirements. The only reason two arrays are required is to allow the permutation to transfer data from the unpermuted array to the permuted array. The permutation information is retrieved from disk in small segments to avoid further depletion of the memory capacity. The memory requirements could be halved by performing the permutation in place. In-place permutation is possible when the permutation matrix is symmetrical, which unfortunately is not the case here. However, it is always possible to factor a permutation matrix into the product of two symmetric permutation matrices [FRASER-76]. For example, the permutation matrices in Eq. (3.13) can be factored as follows:

$$
P_1 = \begin{bmatrix} 1&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0 \\ 0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&1&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&1&0&0&0&0 \end{bmatrix} = \begin{bmatrix} 1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&0&1&0 \end{bmatrix} \begin{bmatrix} 1&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&1&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&1&0&0&0&0 \end{bmatrix}
\tag{3.31}
$$

$$
P_2 = \begin{bmatrix} 1&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&1&0&0 \end{bmatrix} = \begin{bmatrix} 1&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&0&1&0 \\ 0&0&0&0&1&0&0 \end{bmatrix} \begin{bmatrix} 1&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&0&1&0 \end{bmatrix}
\tag{3.32}
$$

The penalty for halving the memory requirements by this technique is that the permutation must be performed in two passes, increasing the computation time.

Beyond this simple modification, it becomes more difficult to reduce the amount of memory required. Algorithms have been developed for performing an FFT off disk [FRASER-79], and it should be straightforward to adapt these to performing the FHT in the same manner. But it will also be necessary to devise a means of performing the permutation off the disk in a reasonably efficient manner. These modifications would make the technique completely general, being able to deal with any quantity of data. Because the amount of storage available on disk is virtually

unlimited, it would always be possible to obtain the maximum advantage in dynamic range from the amount of data collected (Sec. 3.6.3).

It is also intriguing to consider hardware implementation of this crosscorrelation technique. Because of the triviality of the arithmetic operations required, the hardware would be much more straightforward than for the FFT. It is completely within reason to imagine real-time operation of the algorithm using the appropriate hardware.

3.8   CONCLUSION

Measuring the impulse response using a noiselike excitation is capable of providing much greater dynamic range than can be obtained using an impulsive excitation. This advantage is crucial for measuring a concert hall because an impulsive excitation will not overcome the ambient noise. When the noiselike excitation is chosen to be a binary maximal-length sequence, several other advantages accrue. Foremost among these is that the crosscorrelation can be performed efficiently. Because the excitation is binary, only additions are required. Furthermore, an efficient algorithm based on the fast Hadamard transform exists for minimizing the number of additions. A straightforward crosscorrelation would require approximately n^2 multiplications. But this efficient algorithm requires only about $2.5 \log_2 n$ additions, which represents a reduction in the execution time of several orders of magnitude for most values of n. There are also advantages to using maximal-length sequences because they are deterministic signals. Measurements are exactly repeatable so that for time-invariant systems additional improvement in the dynamic range can be obtained by averaging several measurements.

CHAPTER 4

EVALUATING THE MODEL

4.0 INTRODUCTION—OBJECTIVE EVALUATION

Having obtained the impulse response by measuring and simulating the same concert hall, the culmination of the analysis phase is to compare the two. In signal processing, the usual approach would be to compare the two signals on the basis of their mean-square difference. The choice of mean-square error is common because its use is often a sufficient condition for tractable mathematics. But mathematical convenience is not an adequate justification, particularly in this case. In the end, response to a concert hall is subjective, so any objective measure must take into account subjective response. In other words, it is conceivable that the mean-square difference between a simulated and measured impulse response might be large, even though they would sound the same. Although devising an objective measure that incorporates a subjective weighting is conceptually possible, developing the required perceptual model is one of the potential applications of this thesis. Attempting to apply an objective measure before such research is performed would be putting the cart before the horse.

One important illustration of the sort of problem that can arise in objective comparisons

has to do with the directional properties. An objective comparison of the omnidirectional impulse responses would be meaningless because sound fields that differ only in the direction from which sound reaches the listener will still sound different. An abstraction of reverberant fields that has been studied by Barron [BARRON-74] is a single delay feeding two speakers at the same azimuth on either side of the front speaker. He avers that the only subjective effect of delayed sounds when they are laterally displaced is "spatial impression" (SI). The degree of SI is a function of the azimuth of the secondary speakers. Thus, the subjective effect will differ in two sound fields that differ only in the azimuth of the delayed sounds. The omnidirectional response would suppress this effect.

But the problem is even more profound. Were this its extent, the solution would be to use the binaural response because it preserves the directional characteristics. But Barron also found that SI is a function of the amplitude of the delayed sound. Thus, there are an infinite number of ways of generating the same SI by simultaneously varying the azimuth and amplitude of the secondary sounds. Two sound fields might differ objectively and yet produce the same subjective effect. An adequate objective measure would have to take this effect into account.

Real sound fields are considerably more complex than the abstractions studied by Barron. What other subjective effects remain to be discovered in more realistic conditions could only be guessed. Until they are known and quantified it would be impossible for an objective comparison to accurately reflect subjective response.

4.1 SUBJECTIVE EVALUATION

The other approach to assessing the performance of the model bypasses the problem of accounting for subjective response in an objective way by allowing listeners to respond subjectively. A simulation that sounds the same as the actual concert hall is the desired result, even if the mean-square error is large. Unfortunately, subjective comparisons are also problematic: the comparison must be performed in a way that accurately conveys the directional characteristics of the reverberation.

Presenting reverberation monophonically graphically illustrates the distortion that results

when the directional dependence is ignored. One common application of the small portable tape recorders that have proliferated in recent years is for recording lectures or concerts. People are often surprised to find that the recording sounds more reverberant than the live event. The reason is that at the original performance one's auditory system responds to the directions of the different delayed sounds, but the information that makes it possible for us to do this is lost in the monophonic recording.

Stereo is slightly more effective than mono because it allows two reverberance signals (the later reflections) to be presented that have a degree of incoherence, thereby producing a more convincing sense of spaciousness. However, an accurate presentation of reverberation exceeds the capabilities of stereo because sounds must reach the listener from every direction. In stereo, the range of directions is limited to the sector between the two speakers. If all the reflections were recorded stereophonically, most would come from the wrong direction. This distortion can result in the reproduction seeming overly reverberant, just as in mono.

A quadraphonic system, the obvious continuation of this progression, has been used in serious research in artificial reverberation, but it still suffers from inadequate directional resolution. A more extensive array of speakers could be used, but there are some practical limitations. A large number of speakers, power amplifiers, and other equipment necessitates a laborious setup procedure and is expensive. The listening environment must be large and unencumbered to accomodate all the equipment, and ideally it would be anechoic. Also the signal processing is complicated by the need to generate a large number of different signals and route them to the appropriate speakers. A well-known example of this sort of system is the 12-channel system of BBN. More recently, Kleiner [KLEINER-81] reported on a simulation using an array of 50 speakers distributed uniformly along circles at elevations of 0°, 22.5°, 45°, and 67.5°. The University of Göttingen has a system comprising 65 speakers [BLAUERT-83]. Whether 12, 50, or even 65 speakers is enough has not been demonstrated, and yet these systems are already impractical. On the other hand, an array of speakers has the advantage of allowing head movement to be used as an additional directional cue. Several listeners can share a listening experience. With a sufficient number of speakers there would be no question regarding the accuracy with which the directionality was presented—it would be the same as in real life.

A fundamentally different approach is to use a binaural presentation. In binaural record-

ings, the two signals correspond to the sound pressure at the listener's ears. Playback will sound exactly the same as the original performance as long as the sound pressure reproduced at the listener's ears is the same as the recording. Binaural recording is very economical in signal bandwidth insofar as only two channels suffice to present sounds in any direction. This economy derives from the way in which the required directional cues are encoded in the recording. However, this sword cuts both ways: the recorded directional cues correspond to only one listener position. The acoustical effect will not be consistent with any listener movement during playback. In particular, head rotation can not be used to provide an additional localization cue; but head rotation is not essential, so this is not a serious limitation.

There are several basic considerations in selecting a technique for presenting sounds to a listener. One important one is that analyzing the model requires comparing it with measurements obtained in the hall used as the basis for the modeling. Aside from the practical difficulties of dealing with a large number of signals, applying the speaker array to the simulation is very straightforward as all the directional information is available from the model. When using measurements, each speaker would be used to present the sound emanating from the corresponding direction in the actual concert hall. As a result, it would be necessary to measure the impulse response in as many directions as there are speakers, and with a directional resolution corresponding to the placement of the speakers. Making all these measurements not only represents a major inconvenience, but obtaining the required resolution could be a formidable technical challenge as well.

In measuring the acoustics of an existing concert hall, a much more straightforward technique is to record the signal binaurally. Then, only two channels are needed. The recordings are usually made using a dummy head, although real heads can also be used by placing very small microphones into the auditory canal. Having only two channels to deal with also simplifies the mechanics of the simulation, but the signal processing is complicated by the need to simulate binaural hearing. To create the correct impression of direction for each VS, it is necessary to filter its sound with the transmittance to each ear for the appropriate direction. However, measuring the directional transmittance to each ear is much easier than measuring the directional transmittance of a concert hall. Furthermore, the measurement can be performed at the laboratory, and needs to be performed only once. Binaural measurement trades simplicity in the concert hall for complexity in the laboratory.

Another consideration in choosing the method of presentation is that the sounds presented
to the listener must be externalized—that is, they must not sound as if they are emanating
from inside the head. An important advantage of a speaker array is that the sound will almost
certainly be externalized because the speakers are external. Undistorted binaural listening will
also sound externalized, but with headphone listening, obtaining a consistently externalized
image is still difficult probably because of the emplacement of a small resonant chamber over
the ears.

However, a technique first described by Bauer [BAUER-61] and developed by Schroeder
[SCHROEDER-70] makes it possible to audition binaural recordings using speakers. Because
speakers allow free-field listening, the image is more durably externalized than with headphones.
This method circumvents one important objection to binaural recordings, although it is still
true that the listener's position must be constrained, the listening experience cannot be shared,
and the acoustics of the listening room becomes a consideration. On the other side, presentation
with a speaker array has a number of significant disadvantages, principally the difficulty and
inconvenience of measuring the impulse response of existing concert halls in a large number
of directions. The practical difficulties associated with measuring concert halls must receive
a very heavy weight, so the binaural technique would clearly be favored on the basis of these
preliminary considerations.

4.1.1   Reducing the number of speakers

Use of the speaker array still offers a presentation that can be shared and that allows
head movement as advantages over the binaural approach. These advantages make it worth
considering possible simplifications that overcome the practical difficulties before abandoning
the concept. One possible simplification will be considered in more detail in the next chapter:
in less demanding applications one is often willing to ignore the elevational component of
the delayed sound direction because auditory perception tends to be less sensitive to vertical
displacements. Having suppressed the vertical component, analysis by the image model often
reveals that the early delayed sounds originate from a few bands of directions. Because of
the limited directional acuity of hearing, it is reasonable in many cases to approximate the
direction of all the sounds in the same band by presenting them from a speaker positioned at
the average direction. These simplifications make it possible to drastically reduce the number

of speakers—in some cases to as few as four—while still retaining some accuracy.

Unfortunately, these ideas cannot be applied in a scheme that is general and precise. Some concert halls do not display a clustering of directions of the virtual sources. Even in those that do, the directions of the clusters change in different halls. Therefore, comparisons would require that the speakers be moved. Although A-B comparisons could also be performed by setting up two speaker arrays and switching between them, comparisons would still be cumbersome. Furthermore, when one considers the effect of sound diffusion and diffraction, it is no longer true that all the sound arrives from the directions of the VSs. The diffuse sound arriving from other directions might be significant despite its lower amplitude when it is the only sound arriving from those directions. Deciding where to position speakers would require assumptions about the importance of directional cues. Thus, this approach to simplifying the speaker array sacrifices too much generality and makes assumptions that might invalidate the simulation.

Another approach to simplifying the speaker array is to match the placement of the speakers to the directional acuity of auditory perception. Rather than distributing speakers uniformly around the listener, one could increase the density of speakers in regions in which directional acuity is greater, with fewer speakers in regions where hearing is less acute. This matching to auditory perception would apply in the dual situation of measurements in a concert hall, so that fewer measurements would be made in directions less readily distinguished. Enough speakers would be positioned so that increasing the number would not produce an audible change in the simulation. One difficulty with this approach is that the available data in the literature is inadequate for deciding where speakers should be positioned. The well-known study by Mills [MILLS-58] deals only with the directional acuity of isolated sounds. Mills shows that acuity is greatest in front of the listener, and that elevational acuity is lower. Although there is no reason to expect the acuity of a delayed sound heard in conjunction with a direct sound to follow a different pattern, the greater complexity of the situation probably reduces acuity. The elaborate psychoacoustic testing required would be beyond the scope of this thesis.

The more important limitation is that directional acuity is head-related, so a speaker array matched to the characteristics of hearing would require fixing the position of the listener's head. Head rotations would align directions of high acuity with parts of the speaker array with lower than necessary speaker density. Although some head movement could likely still be tolerated,

that is also the case in speaker presentation of binaural sounds. Thus, the primary advantage of the speaker array would be nullified.

A final type of simplification is practiced by Kleiner [KLEINER-82]. To minimize the size of the speakers in the array, he feeds all signals in the lowest octave (100-200 Hz) to two wide-range speakers. He is also able to limit the high frequency response as his work deals only with speech. The size of speakers that are flat only from 0.2–5 kHz can be significantly reduced. He points out [KLEINER-81] that minimizing the size of the speakers is also important to reduce spurious reflections off the speakers themselves.

4.1.2   Simplifying the speaker array

A different approach to simplifying the speaker array is based on the hypothesis that hearing is less sensitive to distortion in the delay channels than in the direct. It is well known that for purposes of determining the direction of a sound source the direct sound is favored. The favoring of the direct sound might extend to other aspects of perception as well. Tolerance to distortion would simplify the speaker array by allowing inferior equipment to be used. For example, in the category of linear distortion, if a loss of high frequency energy were not audible, then speakers with a less extensive high-end could be employed. Even more significantly, if the same were true of the low frequencies, then the woofer, usually the largest part of the speaker system, could be omitted. Although the number of speakers would not change, their complexity would be reduced with a commensurate reduction in their cost and greater ease of positioning. However, the inconvenience of setting up an array of speakers would remain. Also, this approach to simplification would not ameliorate the measurement problem.

The potential advantages of tolerance to nonlinear distortion are harder to see. The numerous power amplifiers could have lower power capacity as occasional overloads would not be audible, but otherwise designing electronic components with inaudible distortion is straightforward. Transducers usually have higher levels of nonlinear distortion, but it is possible that tolerance to still-higher levels would allow simplifications. Any simplifications could reduce the equipment costs somewhat.

Despite the limited potential of this approach, a simple experiment was performed to test
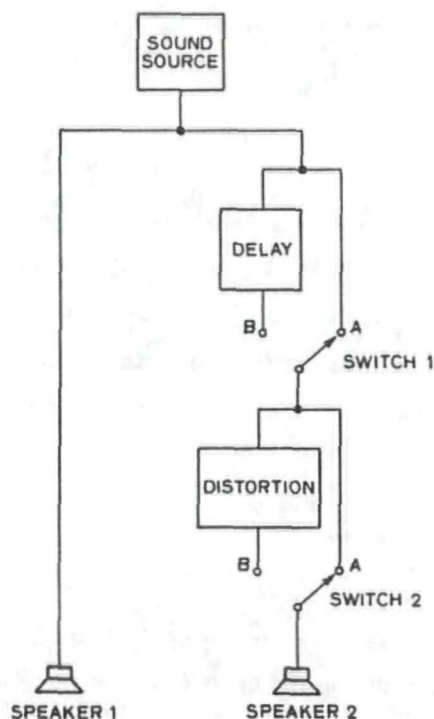
Fig. 4.1 Block diagram of experiment to test the sensitivity of hearing to distortion in a delayed sound heard in conjunction with an undistorted undelayed sound.

this theory (see Fig. 4.1). Speaker 1 is stationary and is positioned directly in front of the listener. Speaker 2 produces the delayed sound; it is moved around the listener at a constant distance to explore the dependence on direction. For any particular position and delay time, the test proceeds in two steps. In step 1, the sensitivity to distortion in the secondary channel with no delay is established. Switch 1 is placed in position A, and switch B is moved back and forth while adjusting the severity of the distortion until it is just noticeable. Now switch 1 is set to the other position, introducing a delay into the secondary channel, and the test is repeated. The hypothesis is that with a delay present in the secondary channel, more distortion will have to be introduced before it is noticeable because the first sound is favored. The experiment makes it possible to investigate the dependence of the effect on the relative delay of the two sounds and the position of the secondary speaker.

Only informal tests were performed as they revealed no change in the sensitivity to the characteristics of the secondary sound whatsoever. The effect of a 10kHz low-pass filter on white noise was clearly audible whether or not the delay was present for delay times of 25 and 50 ms. More careful tests might have revealed some change in sensitivity for cutoff frequencies higher

than 10kHz, but in such a high frequency range the effect would provide little practical benefit. It is also possible that white noise is an unreasonably stringent test. Perhaps with musical signals there would be a reduction in our sensitivity. Despite these caveats there is obviously little hope for a significant simplification of the speaker array. When one also considers that the measurement problem is not ameliorated by this approach, the conclusion is inescapable: the speaker array is not a practical approach.

4.2 BINAURAL PRESENTATION WITH SPEAKERS

Having rejected the use of a speaker array, we return now to binaural presentation. As observed previously, the binaural method offers a significant advantage in measuring concert halls because only two channels need to be recorded. The concert hall simulation is complicated by having to simulate binaural hearing as well, but the required information need be obtained only once, the measurements can be made in the laboratory, and the additional signal processing would be modest compared to the signal processing already required. Reports in the literature indicate that with headphone listening the sound is not consistently externalized [SCHROEDER-73]. But the crosstalk cancellation scheme with speakers [SCHROEDER-70] circumvents this problem by allowing free-field listening. Speaker auditioning brings the acoustics of the listening room into the picture, but moderate sound deadening seems to be adequate to deal with this problem [SAKAMOTO-82]. Otherwise the most serious objection is that the listener is inconvenienced by having his position constrained. Minimizing this constraint is a worthwhile objective.

The basic premise of the crosstalk cancellation scheme is that the difference in position of the two speakers allows the crosstalk signal to be canceled. A heuristic explanation is that a signal is added to the opposite channel that has been delayed, filtered, and inverted so that when it reaches the near ear it cancels the crosstalk from the original channel. Of course, the crosstalk canceling signal also crosstalks, so another cancellation signal must be emitted from the first speaker, and so on. Thus, complete cancellation requires recursive prefiltering before emitting the signals. This prefilter is the inverse of the acoustical transmittance. Details are available in the literature [SCHROEDER-70] [SAKAMOTO-82] [DAMASKE-71].

To implement the technique, we must measure the transmittances to each ear from each speaker for the given positioning of speakers and listener. The transmittance will be different for different listener positions, so any error in the placement of the listener translates into errors in the cancellation achieved. There have been reports in the literature that head rotations of $\pm 10°$ [SCHROEDER-70] and lateral deviations in the listener position of ± 10 cm [SAKAMOTO-82] can be tolerated. Measures must be taken to assure that the listener is positioned with an accuracy within these limits or the binaural effect will be lost. The critical nature of listener placement is inconvenient, uncomfortable, and restricts listening to one person. One should recognize that for signals that are predominantly low frequency, greater inaccuracy in listener positioning can be tolerated. For example, Sakamoto [SAKAMOTO-82] reports that for male speech, the allowable lateral deviation increases to ± 150 cm.

The precision with which the transmittances must be measured also depends on the space in which images are to be positioned. Schroeder and Atal [SCHROEDER-63] synthesized locations only in the front half of the horizontal plane. Morimoto and Ando [MORIMOTO-80] extended the technique to a full three-dimensional space by carefully measuring the transmittance to each ear in a large number of directions. But they considered only individual sound sources. Each virtual source of the concert hall simulation behaves as if it were a distinct source of sound, so the technique must be extended by applying superposition to account for the large number of different directions.

The advantages of this approach dominate no matter how critical listener positioning is. However, one would still like to minimize this sensitivity if possible. One simple consideration is whether the two loudspeakers can be positioned to minimize the sensitivity to position. Schroeder reported using a placement of $\pm 22.5°$, Damaske [DAMASKE-71] used $\pm 36°$, and Sakamoto used $\pm 30°$. These placements seem to be inspired by standard practice in stereo installations—not a compelling rationale from either theory or applicable research. Morimoto and Ando [MORIMOTO-80] positioned the speakers at $\pm 90°$ in azimuth because the head-related transfer function (HRTF) was most uniform for this direction, minimizing computational problems. It is well known that directional acuity is poorest for sound sources directly alongside the listener [MILLS-58], so it is likely that lateral speaker placement also minimizes sensitivity to listener position.

Ideally, movement of the listener during playback would change the sound reaching him in the same way it would have changed at the original performance. Such a change not only makes reproduction more realistic, it can provide additional directional cues. However, movement of the listener during playback can provide cues concerning only the position of the speakers, as the directional cues of the performance are encoded in the binaural recording in a very complicated way. Careful speaker positioning will only make it possible to hear some binaural effect—but not the correct one—for larger positional errors. For example, the sound field will rotate as the listener turns his head. When the positional error becomes excessive, the binaural effect will be lost. It is probably preferable to have the wrong binaural effect than the only alternative, none.

4.3   COMPUTING THE BINAURAL IMPULSE RESPONSE

Previous work in simulating binaural hearing [SAKAMOTO-82] [MORIMOTO-80] has dealt only with individual sounds. The work is extended here by superposing the individual responses for each VS. Starting with a specified geometry, the image model computes the positions of the visible VSs, which is sufficient to determine the arrival time of each delayed sound. The amplitudes can be determined by additionally specifying the absorption coefficients, making it
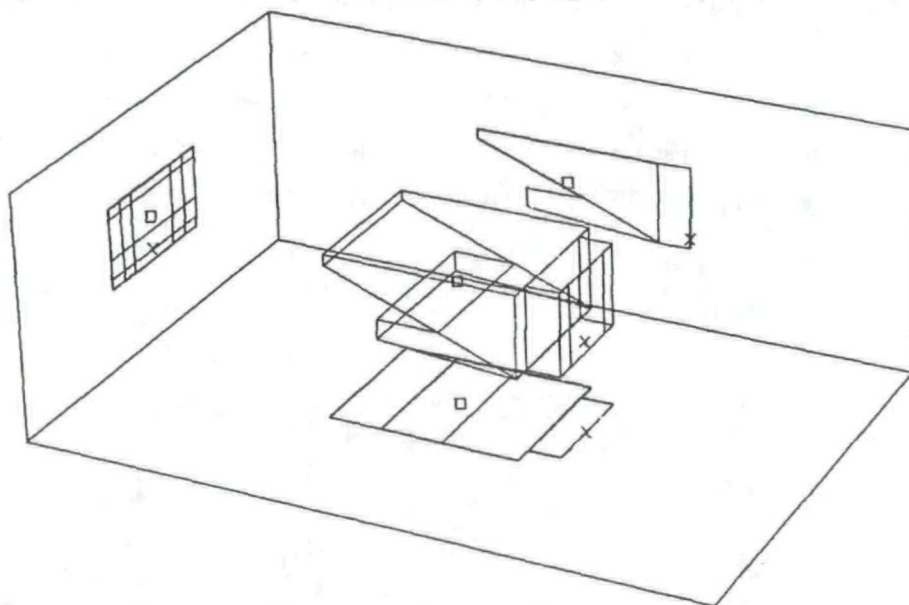
Fig. 4.2   The sequence of operations leading to the computed binaural impulse response begins by describing the geometry of the concert hall.

```
name of output file:  MEM3.imp

number of vs calculated:  226618
number of visible vs:  21
number of vi:  42
maximum level:  4
```

	time	ampl	azim	elev	coords			map			
1	2.648	-83.4	8	-21.	.88	78.88	-27.88	11			
2	2.751	-83.4	8	-23.	.88	69.19	-29.32	12			
3	16.139	-2.8	8	39.	.88	78.88	57.88	15			
4	21.187	-85.5	8	-88.	.88	17.12	-94.44	15	14		
5	22.882	-85.6	8	44.	.88	78.88	67.88	11	15		
6	27.263	-89.1	8	-84.	.88	18.89	-182.26	11	15	14	
7	35.529	-3.8	-51.	-9.	86.88	78.88	-17.88	18			
8	37.258	-87.8	-51.	-14.	86.88	78.88	-27.88	11	18		
9	37.318	-87.8	-51.	-15.	86.88	69.19	-29.32	12	18		
10	38.268	-4.1	52.	-8.	-98.88	78.88	-17.88	17			
11	39.943	-87.2	52.	-13.	-98.88	78.88	-27.88	11	17		
12	48.818	-87.2	52.	-14.	-98.88	69.19	-29.32	12	17		
13	46.588	-4.8	-51.	27.	86.88	78.88	57.88	18	15		
14	49.859	-5.8	52.	27.	-98.88	78.88	57.88	17	15		
15	58.297	-88.1	-79.	-47.	86.88	17.12	-94.44	18	15	14	
16	58.985	-88.1	-51.	31.	86.88	78.88	67.88	11	18	15	
17	52.689	-88.2	79.	-46.	-98.88	17.12	-94.44	17	15	14	
18	53.284	-88.3	52.	38.	-98.88	78.88	67.88	11	17	15	
19	54.891	-11.4	-83.	-58.	86.88	18.89	-182.26	11	18	15	14
20	56.184	-88.5	8	83.	.88	17.12	134.44	15	14	15	
21	57.192	-11.6	83.	-48.	-98.88	18.89	-182.26	11	17	15	14

Fig. 4.3   The image model computes the positions of the VSs. The positions determine the arrival times and, with the addition of absorption coefficients, the amplitudes of all the delayed sounds. The results are summarized in tabular form, with delay time and amplitude relative to the direct sound, azimuth, elevation, Cartesian coordinates, and a map of the sides encountered by the echo.

possible to determine an omnidirectional impulse response. Then, the sound produced by each VS is filtered by the binaural impulse response for the appropriate direction. More precisely, the direction of the VS is used to select the appropriate impulse response for each ear from the set of measurements of the subject. The time origins are shifted according to the distance of the VSs, and the amplitudes are adjusted to account for the attenuation along the corresponding path. These results are summed into the cumulative impulse responses for each ear. Thus, the final binaural impulse response for the concert hall is the result of adding together the partial responses due to the individual VSs. This sequence of operations is illustrated in Fig. 4.2–Fig. 4.5.
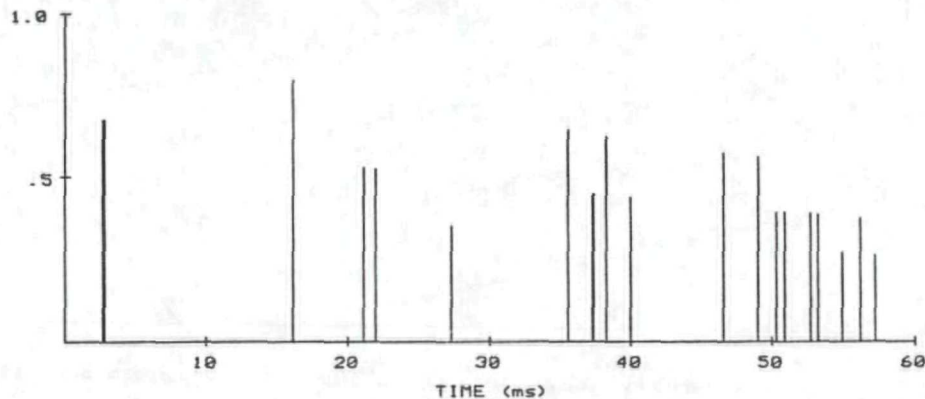
Fig. 4.4   Using the results of the image model, another program computes an omnidirectional impulse response.

chan1

chan2

D:MEMB.BIN
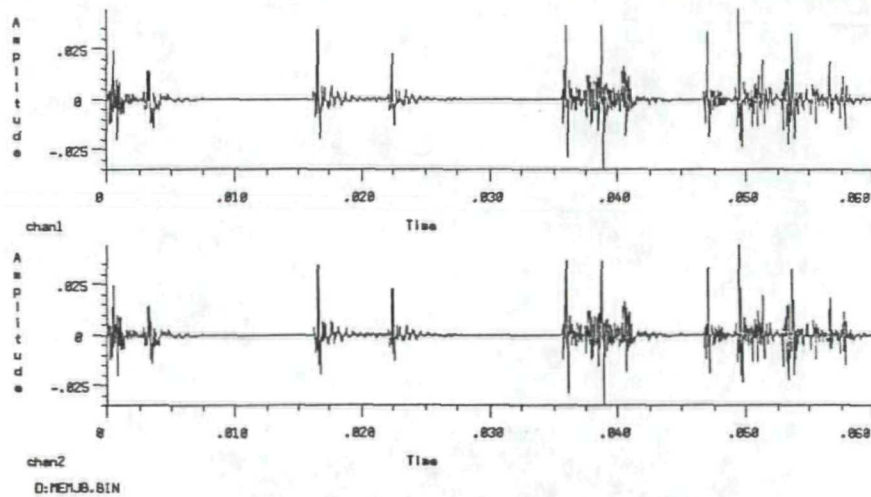
Fig. 4.5   In the final step of the simulation, the direction of each delayed sound is used to select the appropriate measurement of the subject's ears. The binaural impulse response for that direction is added into the cumulative binaural impulse response for the concert hall at the appropriate time and amplitude. The final binaural impulse response presents the reflections to the listener in a way that preserves their directional characteristics.
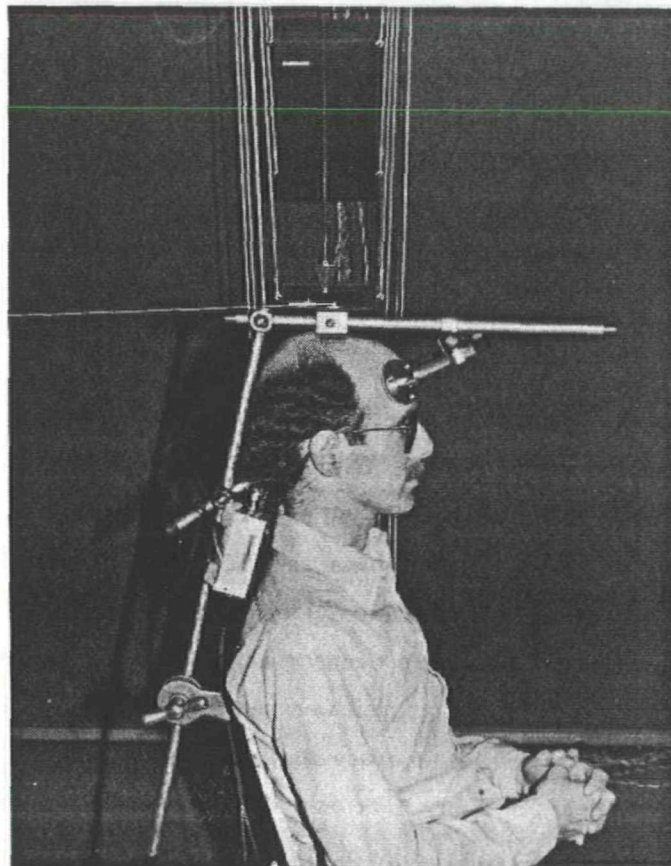
Fig. 4.6   The subject's head must be held in position for the duration of the measurements so that all measurements correspond to the same position. This arrangement provides adequate stability with a minimum of apparatus that could interfere with the sound being measured. Barely visible are the microphones in the subject's ears at the opening of the auditory canal.
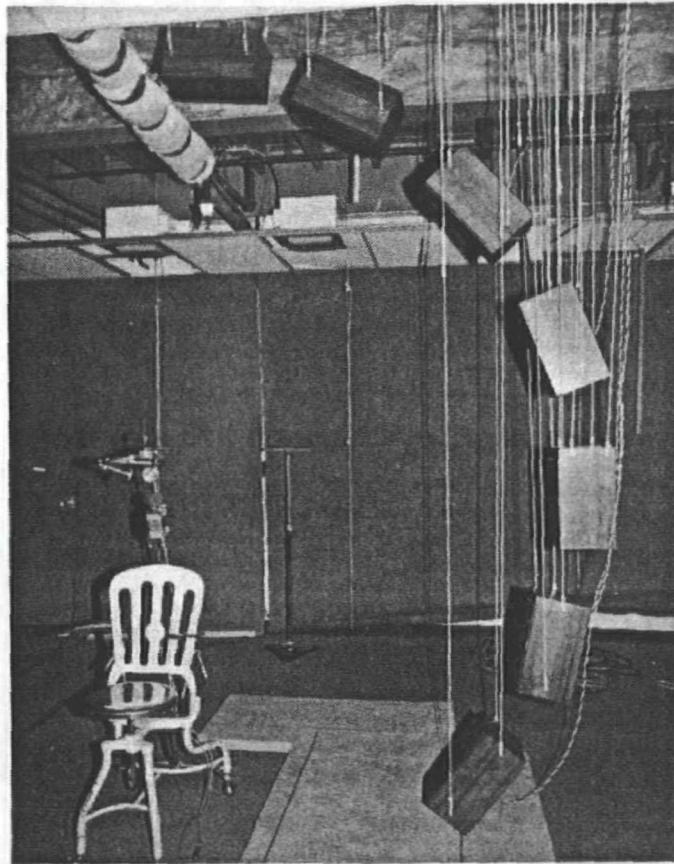
Fig. 4.7   By positioning speakers along a vertical arc, the elevation of a measurement can be controlled by selecting the appropriate speaker. Rotating the subject changes the azimuth.

The directions of the VSs must be quantized. One of the attractions of binaural presentation is that fine quantization is inexpensive. With speakers, fine quantization not only requires that many speakers be used, but the concert hall must be measured in many directions. With binaural presentations we have the dual problem: the HRTF must be measured in many directions. But these measurements are easier because they can be done in the laboratory, and they need be done only once (rather than once per concert hall). Ultimately we would like to whittle the set of directions down to the fewest possible to expedite the measurements and simplify the processing, but the convenience of this approach allows us to start with a large set rather than make assumptions that might prove false. The only limitations to the number of directions are the patience of the experimenters and the size of the computer memory. The size of the speaker can also be a limitation because it would be meaningless for the resolution of the measurements to be finer than the angle subtended by the speaker. In principle the speaker could be moved farther away, but any indoor measurement space will eventually impose some limit.
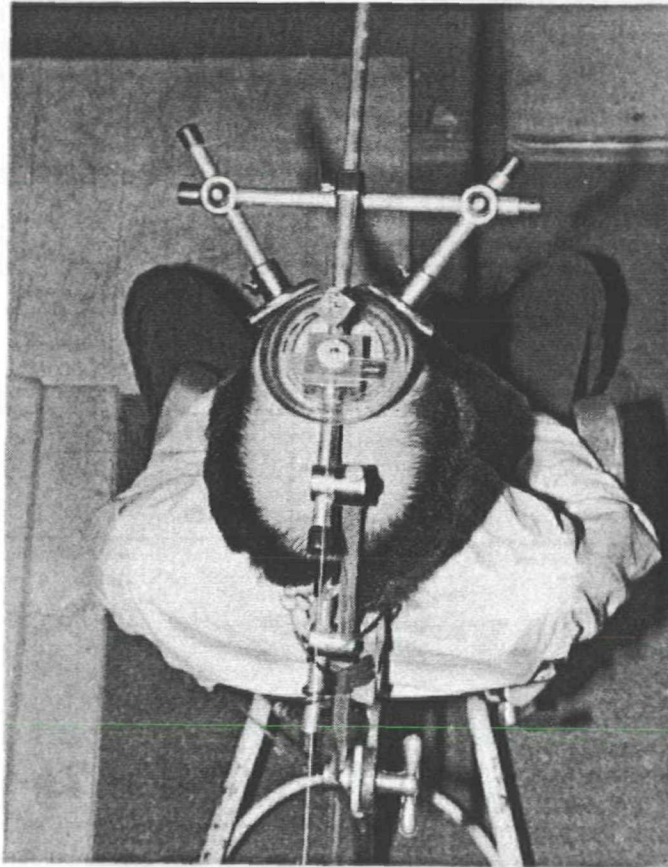
Fig. 4.8   A simple technique was used to measure the azimuth: a protractor was positioned over the subject's head centered on the axis of rotation. An elastic thread attached at the rotational axis and anchored at the other end served as a cursor. The plumb marked the axis so the subject could be easily realigned after each rotation.

The need to keep the subject constrained for the duration of the measurements (Fig. 4.6) requires an efficient procedure. Having to reposition a speaker at different azimuths and elevations would be too time-consuming. One solution is to position several speakers along a vertical arc of a sphere (Fig. 4.7). The elevation of the measurement is controlled by selecting the appropriate speaker, with different azimuths obtained by rotating the subject (Fig. 4.8). Azimuth was measured every 10°, and elevation at 20° increments from −40° to +80°. Lower elevations were considered unimportant because of the high attenuation of sounds reaching a listener from these directions in actual concert halls. These spacings were a compromise between practicality and auditory acuity. Morimoto and Ando [MORIMOTO-80] measured localization errors for real sounds in the horizontal plane of 7–9° and in the median plane of 10–19°. The choices of 10° resolution in azimuth and 20° in elevation are just outside their ranges. There is no point in simulating directionality more accurately than can be perceived. These choices
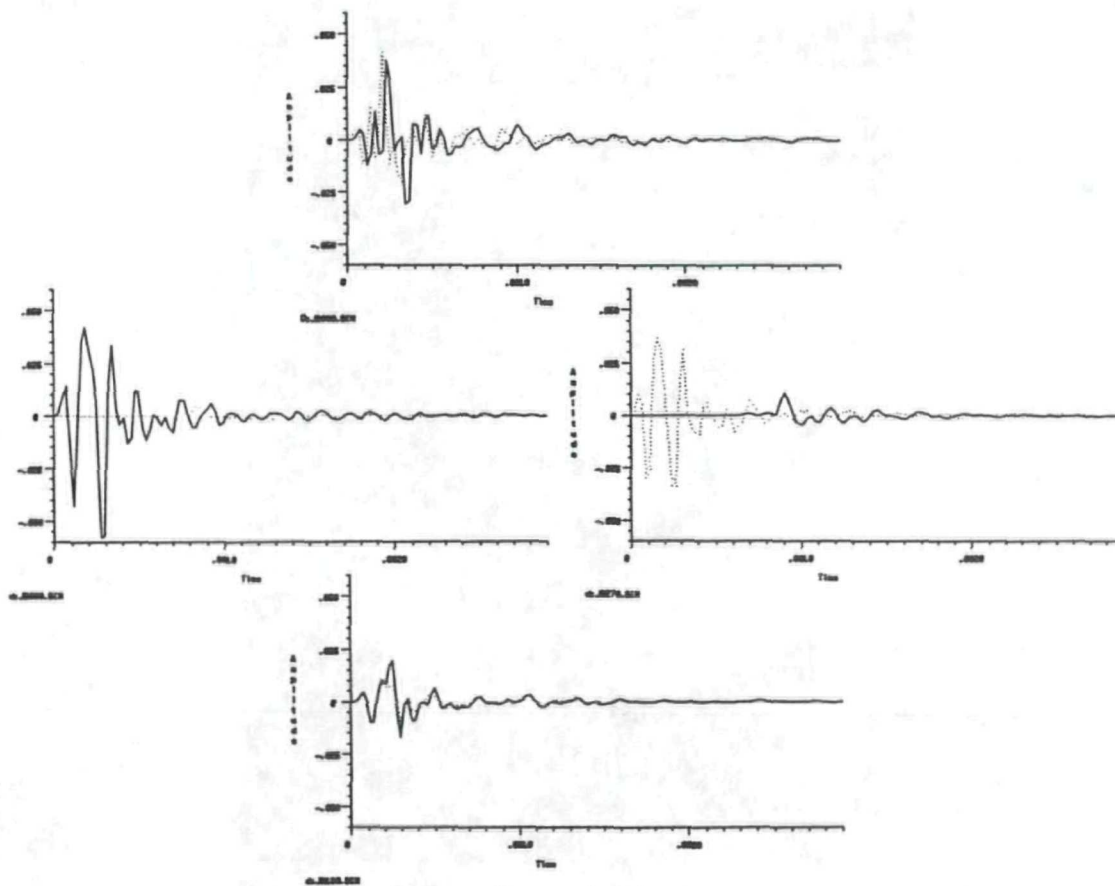
Fig. 4.9 Some examples of the binaural impulse response. The measurements shown were all made at 0° elevation and at 0°, 90°, 180°, and 270° in azimuth.

result in a set of measurements at 36 azimuths and 7 elevations for a total of 252 binaural measurements (Fig. 4.9). Basing the presentation on binaural hearing makes simulating this many directions feasible, whereas an array of 252 speakers would clearly be unreasonable.

Following the practice of Morimoto and Ando, the speakers were positioned 1.5 m away from the subject. We used Boston Acoustics A40 speakers, chosen for their small size, low price, and surprisingly flat response (within ±3 db from 67 Hz–20 kHz). At this distance, the speakers subtend 6°—obviously small enough not to obscure the measurement in the horizontal plane. However, the spacing decreases away from the horizontal plane according to $\cos\theta$, where θ is the angle relative to the horizontal plane, so the measurements are just overlapping for the ring at 60° elevation. The procedure was sufficiently convenient that it was not considered necessary to sample the measurements for 80° elevation despite their overlapping.

4.4   CONCLUSION

Validating any model for a concert hall requires a comparison with reality. The comparison is best performed subjectively because of the difficulty of objectifying subjective response. Even then the comparison is not straightforward because the directionality of the sounds must be accurately conveyed. The best solution is to base the presentations on the binaural impulse response. That simplifies measurements in the concert hall because only two signals need to be recorded. The concert hall simulation must include a simulation of binaural hearing, requiring that the sound produced by each VS be filtered by the measured binaural impulse response of the subject for the appropriate direction. Detecting differences will indicate that improvements are necessary and possibly suggest what they are.

CHAPTER 5

SIMPLIFICATION FOR DOMESTIC USE

5.0 INTRODUCTION

In the previous chapter, we considered techniques for presenting reverberant sound as accurately as possible. Basing the presentation on binaural hearing makes it possible to take account of the directional dependence of the sounds with any desired degree of precision. However, the precision is not without cost—the technique is laborious to apply and requires sophisticated equipment. One application that permits less precision is creating a reverberant effect at home. Domestic audio reproduction suffers because it omits many of the reflections we hear in real life. Reproducing these reflections has more than academic significance: usually a significant portion of our enjoyment of a musical performance is attributable to the acoustics of the listening space. Unfortunately, accurately presenting these reflections in stereo is impossible because they reach the listener from directions outside the range stereo is capable of presenting. It is well known that in stereo the range is limited to the sector between the two speakers. This limitation of stereo suppresses one desirable aspect of the listening experience.

The problem we face then in trying to increase the realism of audio reproduction is that the

reflected sounds must be presented, but their range of azimuth is greater than can be presented by stereo. To deal with this problem, three different approaches are conceivable. The first is to record the sound binaurally. This approach has the important advantage that only two channels are required, so that the same audio technology as is currently used for stereophonic recordings could be applied. But elevating binaural recording to a commercial scale would be impractical. To begin with, many commercial recordings are not of live performances, particularly in the case of popular music where the musicians might not even have been in the same room at the same time. Often recordings are made by mixing together many separate signals. It would be difficult to accomodate traditional recording methods within a binaural framework. Another important disadvantage of binaural recordings is that in general, playback is restricted to headphones. The constraint imposed on the listener's position by the cross-talk cancellation scheme [SCHROEDER-70] [BAUER-61] would be unacceptable outside a research environment. Although compromises embodied in the rare commercial products still overcome the directional limitation of stereo without rigidly constraining the listener, they probably are not accurate enough to present the early reflections. Another issue is that binaural recording assumes the recorded signals match the sound that would have existed at the listener's ears had he been present at the original performance. However, differences between individuals make it impossible to satisfy this assumption universally. A final problem with the binaural approach is that performances must be recorded in a special way; salvaging existing stereophonic recordings is impossible.

A second approach to fully recapturing an auditory experience is to use additional signal channels. By positioning a number of speakers around the listening room and driving each speaker with its own signal, it would be possible to accurately present the direction of incidence for each sound. This approach has the advantage of allowing the listener greater freedom than binaural presentation, but it can have formidable practical limitations when the number of additional channels is large. First there is the inconvenience and expense of having to position a large number of speakers around the listening room. A new technology would be required to record and broadcast the multiple channels discretely. The recording process would be complicated by the need to obtain several directional recordings of the reverberation. But again the most severe limitation would be incompatibility: the vast library of existing stereo recordings could not be utilized. New issues would have to be released in two formats, requiring a double inventory; and special playback equipment would be needed for the multichannel

recordings.

Some of the limitations of multichannel recordings apply mainly when the reproduction requires a large number of additional channels. We will see in Sec. 5.1.3 that only two additional speakers are required for an adequate presentation. These more modest requirements bring to mind the quadraphonic systems which existed several years ago. Although they also used four speakers, quadraphonic systems differed in important ways from the system envisioned here. The matrixing approaches, such as SQ, neatly circumvented the compatibility problem because the same record could be used for one-, two-, or four-channel reproduction depending on the playback equipment. Furthermore, the same system could be used for broadcasts and tape recordings. Unfortunately, the techniques that attempted to encode four channels of sound into only two were severely limited in the degree to which the original four channels could be recovered. These quadraphonic recordings created a spatial effect by introducing sounds that were partially uncorrelated with the original stereo information. In fact, the superiority of stereo over mono has more to do with that same spatial effect than with the ability to distinguish the positions of individual performers. Although many found the ambience effect pleasing in some cases, accurate reproduction of auditorium acoustics requires the presentation of sounds that have been delayed significantly. These sounds must be recorded discretely, a capability the matrixing techniques lacked. The CD-4 system used frequency division multiplexing to record the rear channels in a higher frequency band with greater separation between the four channels. It could theoretically have been used to solve the ambience problem, but only if different signals had been recorded in the additional channels and if the speakers had been positioned differently than was the standard practice. Also, this system suffered from incompatibility as well as many other practical difficulties. So in the end, only the third approach to providing ambience furnishes the needed capabilities while preserving compatibility.

The third approach is to simulate the acoustics of a performance space electronically. By applying our understanding of how an auditorium affects a sound produced within it, we can process a signal electronically to create the same effect. Recording the reflections is not necessary because they will be synthesized during playback. Therefore, this is the only approach we have considered that does not require special recording or encoding techniques. It is worth mentioning that the technique is also compatible with monophonic recordings. With this approach, reproduction need not be restricted to the acoustics of the auditorium the music

was actually performed in; it is equally effective at recreating the acoustics of any hall. Thus, another important advantage is that simulation gives the listener control over the environment in which he would like to hear the music, rather than allowing the recording engineer to arrogate responsibility. The principal limitation of simulation in any situation is that the model it is based on is inevitably a simplified representation of reality. However, the inaccuracies of the model need not invalidate the approach if the simplifications are made with an understanding of the subjective implications. The implementation presented in this chapter is capable of substantially increasing the realism of audio reproduction.

The simple simulation also begins by modeling auditoriums with the techniques presented in Ch. 2. The derivation deviates in the way the directional dependence is presented to a listener. We begin the presentation at this point. Several more subtle questions are deferred until the penultimate section, and the chapter concludes with a few comments concerning practice.

5.1   DETERMINING THE DIRECTIONAL IMPULSE RESPONSE

5.1.1   Theory

The virtual source positions computed by the image model characterize the most important acoustical properties of the concert hall because they are a way of describing the paths along which sound travels. The second step of the simulation process is to reduce this information to the directional impulse response. Electronically convolving an audio signal with this impulse response will produce the same effect as if the sound had been emitted in the auditorium. The first steps of this process—determining the delay times and amplitudes of the reflections—were described in Ch. 2.

The simplicity of the domestic simulator results from the way in which the directional dependence is accounted for. Binaural presentation is not an option because most consumers object to headphones and the speaker method is too restrictive. As discussed in Ch. 2, the most precise presentation with speakers is obtained by positioning speakers around the periphery of the room in directions corresponding to the directions of the virtual sources. Unfortunately, the mammoth speaker array required is not practical. The speaker array must be simplified,

but before we can discuss simplifications, we must deal with the basic question of truncating the impulse response.

5.1.2   Truncation time

The amplitude of sound from remote parts of the virtual source lattice is negligible because of the distance the sound travels and the absorption it encounters. Consequently, the impulse response can be truncated. One logical choice is to truncate the impulse response at the reverberation time, defined as the time required for the sound pressure to decay 60 dB. However, most halls have thousands of reflections within their reverberation time. Dealing with all these reflections would require a signal processor with extraordinary speed as well as possibly requiring an extravagant number of speakers.

As discussed in Ch. 1, the impulse response can be truncated much earlier because of the two-phase nature of reverberation. The high density of the later reflections imparts a noise-like quality to the reverberance. Therefore, reverberance can be recorded adequately because, aside from obvious issues of fidelity, it is only its incoherence that must be preserved. If an attempt were made to synthesize the reverberance, then to an extent one would be duplicating a signal already present. In contradistinction, the ambience is largely excluded from recordings because the direction of the reflections is critical. Consequently, the approach taken is to rely on the reverberance present in recordings, and regenerate only the information that must be excluded, the ambience. Truncating the impulse response at the transition between the ambience and the reverberance—usually between 100–150 ms—reduces the number of reflections to 100 or more. That number is still large, but it can be further reduced by recognizing properties of auditory perception. In addition to preserving compatibility with existing stereo recordings, this solution balances the capabilities of the two technologies. The recording brings as much information to the reproduction as it can, and the simulator synthesizes no more than the information that cannot be easily recorded.

5.1.3   Minimizing the speaker array

The basic thrust of the initial simplifications is to reduce the number of speakers required to adequately present the directionality of the early reflections. The first approximation ignores
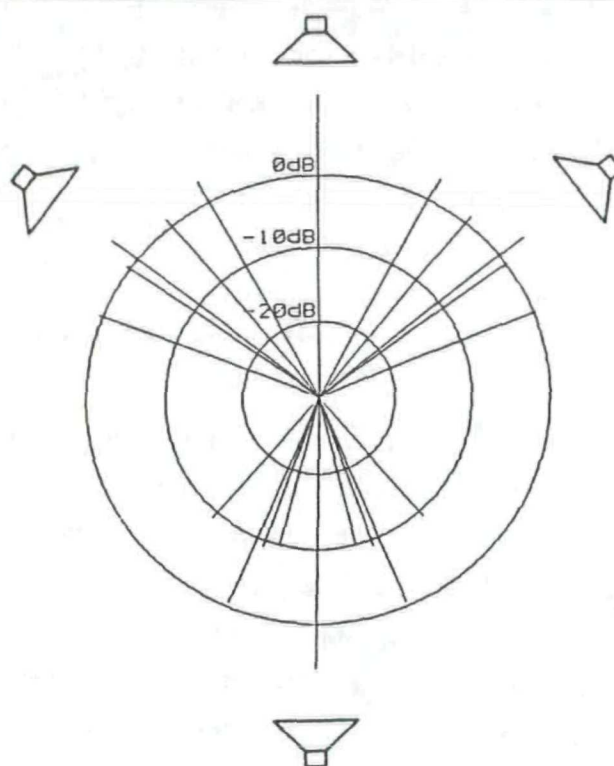
Fig. 5.1 A polar plot of sound power as a function of azimuth shows a characteristic grouping of the directions. All the sounds in each band can be presented with a single loudspeaker positioned at the average direction.

the elevational component of the virtual source positions. The rationale for this step is that our hearing is less precise in assessing a vertical displacement of a sound than a lateral displacement [MORIMOTO-80]. A polar plot (Fig. 5.1) becomes the appropriate presentation to show where speakers must be positioned. Placing speakers at each of the directions indicated in the polar plot would still require an excessive number, but we can take advantage of the fact that the sounds group into bands. Because of the limited directional acuity of our hearing, it is reasonable to present all the sounds within each band with a single speaker positioned at the average direction. Using this approach, the concert hall model considered in this display requires only four additional speakers to present the reflections. The stereo speakers are retained to preserve the existing localization of the performers.

Requiring six speakers would still be an imposition for most listeners. But by carefully analyzing the objective, a further simplification to only four speakers can be made. The goal of a simulator for domestic use is to provide several sets of parameters representing a spectrum of desirable listening environments. Although the simulations must be based on accurate modeling,

there is no reason for the basis of the model to be any particular existing hall. A more reasonable objective is to provide listeners with positions corresponding generally to good auditoriums, even if the simulation does not correspond to any in particular. In fact, there is no reason the model must correspond to any hall that could or would ever be constructed. To the extent we understand what makes some listening environments desirable for certain types of music, electronic simulation allows us to improve on existing halls.

Recent psychoacoustic experimentation [BARRON-71] has found that delayed sounds create a subjectively desirable effect called spatial impression (SI) only when they are displaced laterally from the direct sound. The principal subjective effect of vertical displacements is spectral coloration due to comb filtering, and a slight image shift. Architectural acousticians have begun to consider some of these ideas in designs of new concert halls. For example, Harris, in the design of Avery Fisher Hall, chose a configuration for the ceiling intended to deflect more of the sound to the sides [BLIVEN-76]. Schroeder [SCHROEDER-79a] presented a theoretical analysis of a ceiling with highly diffusing characteristics with the same goal in mind. The successful designs of the past usually have a high degree of SI, and the goal of contemporary designs is to increase the degree of SI. The models available in an auditorium simulator ought to pursue the same objective.

The preeminence of lateral reflections lends additional weight to the notion of disregarding the elevational component of the delayed sound direction. To the extent the elevation can be detected, it does not increase the desirability of the effect, so it is ignored. But now we go further and observe that the front and rear speakers are used to present reflections that are also not laterally displaced. As there is no positive subjective effect associated with these reflections, the center speakers are simply omitted, further simplifying the system. In fact, our experiments have shown that these speakers are otiose. Only by careful auditioning with select musical passages is it possible to tell whether or not the center speakers are active. Requiring only two additional speakers is a practical solution that still achieves a high degree of realism.

No single speaker placement will be appropriate for every concert hall geometry. The polar plot in Fig. 5.1 is based on an analysis of a basically shoebox-shaped hall. Different halls will produce different polar patterns, but fortunately, as far as the important lateral bands are concerned, their azimuth usually does not deviate greatly. This is shown in Fig. 5.2 where
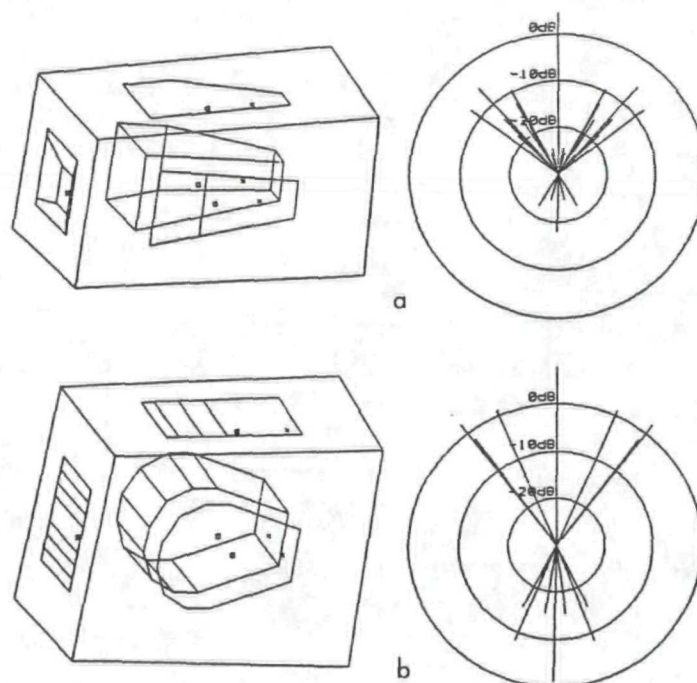
Fig. 5.2 The polar plots for a fan-shaped hall (a) and a horseshoe shape (b) show bands in approximately the same directions as for the rectangular hall in Fig. 5.1, so the same speaker placement can be used in every case.

the polar distribution for two other geometries can be seen to lead to approximately the same speaker array.

There are additional simplifications crucial to reducing the cost of the simulator. Because these simplifications are related to the hardware design, they will be considered in conjunction with the discussion of the signal processor.

5.2 THE SIGNAL PROCESSOR

5.2.1 Basic requirements

The results of the image model can be presented in terms of the impulse response of the auditorium for the given source and listener positions. To electronically synthesize the ambience of the hall, an audio signal must be convolved with this impulse response. A more intuitive way to think of this convolution operation is to consider the signal processor to be a delay line with
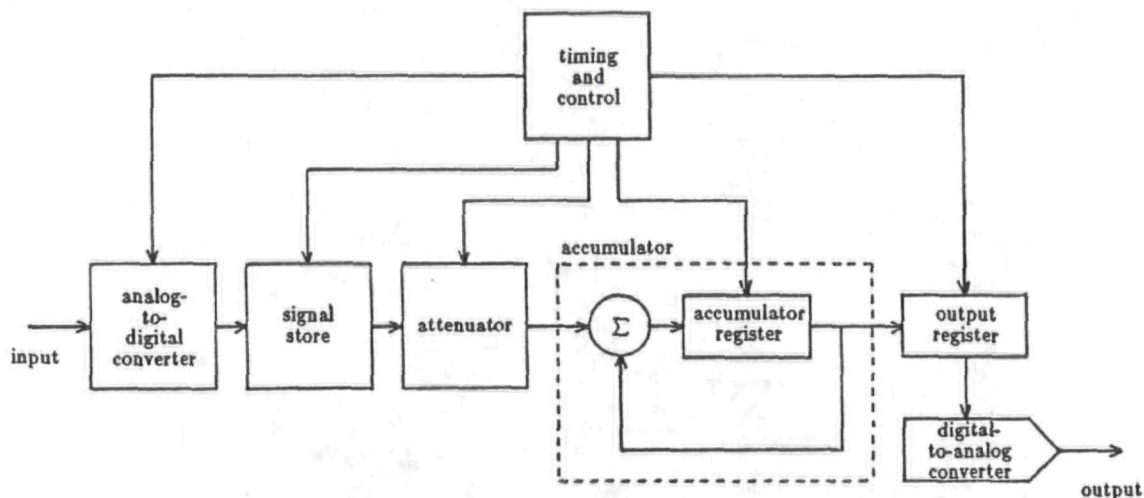
Fig. 5.3   The block diagram of the signal processor.

multiple taps at times corresponding to the times of arrival of the delayed sounds. The output of each tap is attenuated to take account of the sound attenuation along the corresponding path. Then all the attenuator outputs are summed together to form the signal that drives one of the secondary speakers. This process is the direct form implementation of a finite impulse response (FIR) filter, usually known as a transversal filter.

The transversal filter is implemented using digital technology because it would be difficult to maintain the fidelity of the audio signal otherwise. The block diagram for the system is shown in Fig. 5.3. After the analog-to-digital conversion, the signal is stored in the signal store (SS). The SS is based on dynamic RAMs rather than the more conventional shift registers to facilitate changing the model. Consequently, the SS must be accessed sequentially. The SS is configured as a circular buffer [BORISH-83]. In this scheme, the signal stored in the SS remains stationary in the physical memory, while the beginning of the buffer changes. The location relative to the zero-delay pointer at which the SS is accessed determines the delay time of the tap. Refreshing of the dynamic RAMs occurs automatically as a consequence of their being configured as a circular buffer. Each tap output passes through the attenuator into the accumulator. An accumulator is necessary rather than a summer because the taps are processed sequentially. The accumulator is reset once at the beginning of each accumulation cycle. The final accumulated value is passed on to the output register whose output is converted back to analog by the DAC.

5.2.2 Simplifications to the signal processor

A great deal of the expense of the processor is related to the number of taps that must be processed. The amount of time available to process all the taps is limited by the sample period, which is related by the sampling theorem to the desired signal bandwidth. Because of the sequential nature of the processing, requiring more taps reduces the amount of time available to process each one, increasing the expense of the circuitry. However, some finesse is required to reduce the number of taps without destroying the quality of the synthetic ambience.

The most important simplification is based on the limited temporal acuity of hearing. Many of the reflections identified by the image model are so closely spaced that they cannot be resolved by our hearing. Accordingly, there is no need to process them separately. Research with isolated sounds [GREEN-71] indicates that the value for temporal auditory acuity is about 1 to 2 ms. One would not expect hearing to be more acute in the more complex situation of a sequence of delayed replicas. Therefore, reflections more closely spaced than 1 ms can be combined, their powers adding, without producing an audible change in the simulation. Note that the resolution of delay time after combining is still one sample period. The inaudibility of combining was demonstrated using a special high-performance ambience synthesizer developed for experimental purposes that was capable of generating a very large number of taps. Switching between simulations before and after combining revealed no audible difference until the combining range extended to about 5 ms.

Another important question has to do with how the two stereo channels are handled by the simulator. One possibility is to sum the two channels before the signal processing. Then the ambience corresponds to assuming the sound source to be a point midway between the two channels. The other possibility is to assume that the two stereo channels correspond to two point sources emitting independent signals. The signal processor would have to treat the two signals separately, essentially doubling the hardware requirements. Because musical sources actually have some finite extent, one would expect a two-point approximation to sound better. The high-performance simulator allowed us to compare simulations based on these two assumptions. We found the difference very subtle unless the spacing between the two point sources was exaggerated. That produced an exaggerated stereo effect that was distinctly less realistic. The explanation for this unexpected finding is probably related to the fact that in

stereo the two channels are *not* independent. If they were, this approach might sound more realistic than the monophonic system, but as it is, we are apparently no worse off processing a monophonic signal.

Having restricted the source to be symmetrically positioned midway between the two speakers, if we also restrict the listener to be symmetrically positioned in a symmetrical hall, the signals required by the secondary speakers are identical. Rather than a separate set of taps for each secondary speaker, we can generate one ambience signal and use it to drive both speakers. Performing spaces are almost always symmetrical, and listeners almost always attempt to position themselves symmetrically within them, so this restriction would not be objectionable for most cases. This step halves the number of taps required. Barron [BARRON-71] reports that, in an anechoic environment, having identical sounds reach the listener from the sides is a degenerate situation that reduces the spatial impression. However, this problem will not arise in a domestic environment. Not only are there incidental asymmetries arising from the furnishings, the placement of the speakers, and the shape of the room, but the separate stereo channels preserve a degree of incoherence. By assuming the concert hall is symmetric and by combining closely-spaced reflections, we can effectively deal with over 100 reflections, while processing only around 15 taps. This dramatic simplification is pivotal in reducing the cost of the hardware.

Another operation required by a transversal filter that is traditionally very expensive is the attenuation. However, we based our attenuator on the well-known fact that the perception of loudness is basically logarithmic in character. Jesteadt, et.al. [JESTEADT-77], give figures in the range of 1 dB for the minimum perceptible intensity change. However, their study was based on isolated pulsed sinusoids. In practice, the minimum audible intensity difference is usually worse, particularly in the much more complicated case of a sequence of delayed sounds heard in conjunction with a direct sound. Using the special high-performance processor, we were somewhat surprised to find that the difference between a simulation with a 3 dB resolution of tap transmittance and one with only a 6 dB resolution was almost imperceptible. 6 dB gain steps can be very easily obtained in the digital domain by simply shifting the signal, a trivial operation.

Additional savings in the hardware required were obtained by following two general design

principles. First, extensive pipelining was employed. Pipelining is a technique that is common in computer design for increasing the amount of simultaneous computation. Because each successive stage of the system requires the result of the previous stage before beginning its operation, requiring that all the processing be completed in a single processing period would result in each stage being active only part of the time. But by storing intermediate results, each function block can be allocated an entire processing period to complete its operation. Of course that means that the total computation time is greater, but because the simulation is based on delay anyway, this small additional pipeline delay is an insignificant error. It would even be possible to adjust the tap delays to compensate for the pipeline delay because the delay of the first tap is always much greater.

The other important principle was the use of serialism. Thanks to all the simplifications discussed previously and especially the use of pipelining, the speed of operation required to perform some functions is much less than the speed of the circuit components employed. Rather than allowing the circuit to sit idle part of the time, the capabilities of the components can be better utilized by breaking up the operation into smaller pieces and assembling the final result from sequential computations. The extent to which the operation can be subdivided varies from one part of the circuit to another. It was determined by the speed of the processing required, and by the configuration of the standard logic components employed.

5.3   SECONDARY CONSIDERATIONS

5.3.1   Effect of speaker position

Although only two additional speakers are required in the auditorium simulation, a related concern is how critical the speaker positioning is. The ideal placement for the secondary speakers is to the sides and slightly in front of the listening position, because that corresponds to the directions from which sounds would actually reach the listener in the halls. Unfortunately, not everybody will be able to place speakers at these ideal positions. Our tests indicate that in practice, the speakers can be positioned within a reasonably wide range of directions without impairing the effect. Our findings are supported by Barron [BARRON-81] who found that the degree of SI is a function of the position of the speakers, but the sensitivity is low for lateral

directions. In most rooms, we have had acceptable results with the speakers positioned at 45°
to as much as 135° on either side of the listener. Also, the speakers need not be symmetrically
positioned.

A more unexpected discovery is that an acceptable effect can be obtained using only one
secondary speaker. This finding is also supported by the research performed by Barron, who
used two symmetrically placed secondary speakers only to avoid an image shift which interfered
with the A-B testing he was performing [BARRON-81]. Eliminating one secondary speaker is
equivalent to covering the corresponding wall in the simulated hall with absorbing material. As
a result, the change in the simulation corresponds to a real change in the hall being simulated,
and does not introduce any artificial aspect to the simulation. This unusual flexibility in speaker
positioning will facilitate their accomodation in home listening rooms.

Many listeners either by choice or necessity have their stereo speakers positioned at angles
so much greater than the usual practice of ±30° that they enter the range that is appropriate for
the ambience speakers. In such cases one might suppose that the ambience signal could be fed
into the stereo speakers thereby reducing the system complexity even further. Although some
spatial impression is still created in this arrangement, the effect is distinctly less realistic. The
ear is no longer able to separate the ambience from the direct sound as they are both coming
from the same direction. The result is a combfiltering that is an objectionable coloration. The
same problem will occur even when the ambience is fed to separate speakers if the speakers are
positioned near the stereo speakers. For the best effect, the secondary speakers must be outside
the sector of the stereo speakers.

5.3.2 Effect of listener position

The simulation is also relatively insensitive to the position of the listener within his listening
room. Movement of the listener introduces errors in the delay times of the synthesized echoes,
but compared to the nominal delay times, the errors are small. In fact, the delay times
change approximately the same way they would for a corresponding movement in the real
hall. Unfortunately the same cannot be said for loudness, but the only serious problem occurs
when the listener moves into close proximity of one of the speakers. Then the sound from that
speaker will dominate his perception. As long as the speakers are reasonably well balanced, the

listener will properly perceive the ambience regardless of his position in the listening room.

5.3.3   Effect of listening room acoustics

The auditorium simulator has been tested in many different listening environments, and has been found to perform adequately in every case. Not surprisingly, the greatest realism is obtained in fairly dry rooms. However, the degree of dryness need be no greater than that achieved by typical home furnishings.

Live listening rooms will generate delayed replicas of the original signal acoustically, paralleling to an extent the electronic processing of the simulator. However, it must be emphasized that no matter how live it is, the room will not create the same effect as the simulator because its echoes have delay times that are too short and too centralized. Also, changing whatever effect the room *does* create would be inconvenient. Several speakers have been designed to maximize the effectiveness of the natural acoustics of the listening room by projecting part of their output in different directions. By this means they succeed in increasing the degree of lateralization, but overcoming the limitations in the delay time is impossible. Although many find their effect satisfying, the effectiveness of these speakers is limited by the acoustics of the room, the influence of the furnishings, and the position of the speakers. An electronic simulator provides additional degrees of freedom that make it possible to create reflection patterns much closer to those encountered in actual halls.

5.3.4   Effect of recorded reverberation

Although efforts are made to minimize the amount of ambience recorded, it is inevitable that some of the ambience of the original performance will be recorded along with the direct sound. It must be emphasized that no matter how the ambience is recorded, it cannot create the same effect as the ambience synthesized by the auditorium simulator. When the recorded ambience is emitted by the stereo speakers alone, little spatial impression is created because the echoes come from the same direction as the direct sound. The primary subjective effect of reflections in such a case is spectral coloration [BARRON-71]. The creation of spatial impression requires that the echoes originate from a direction outside the range of the stereo speakers. In a real hall, these lateral echoes are basically delayed replicas of the direct sound. But when

the stereo signal is processed by the simulator, the electronically delayed replicas consist not only of the direct sound, but the recorded ambience as well. Therefore, one might reasonably wonder what effect the recorded ambience will have on the simulation.

To deal with this question, the nature of the recorded ambience must be understood. In the recording, any strong echoes received after a significant delay would produce an objectionable spectral coloration due to comb filtering. Consequently, the strong sounds must be restricted to a short interval after the direct sound. This early ambience could be better described as a broadening of the direct sound because there generally are not distinct echoes. This broadening has an effect on the timbre, but does not produce any of the other subjective effects of reflections. Broadening of the direct sound is a natural characteristic of real world listening that arises from several causes. Because musicians never perform in total isolation, the sound they produce bounces off nearby surfaces such as music stands, the floor, chairs, other musicians, and so on. The large number of such reflectors creates a virtual continuum of early reflections. The effectiveness of these reflectors varies with frequency, introducing dispersion. Furthermore, there is also a problem of defining exactly what is meant by direct sound. Instruments are not point sources. A grand piano has a large sounding board. Bass violins and cellos are acoustically coupled to the stage, providing an even larger sounding board. The sound from different parts of the instrument arrives at the microphone at different times, but all within a short interval after the first sound. The cumulative effect is that the recorded ambience amounts primarily to a broadening of the direct sound.

Because the ambience is synthesized from the direct sound, there is also a broadening of each reflection in the ambience. This broadening is also natural, arising from many of the same causes as for the direct sound as well as physical processes that have been neglected in the image model. For example, the image model assumes that reflections are specular, and that the reflection coefficient is independent of frequency. Accurately modeling these effects would produce the same sort of broadening of the impulses in the response of the secondary channels as is created by the recorded ambience. This is another example of how the recorded material contributes to a more realistic simulation.

The explanation of how recording engineers restrict the ambience to short delays differs for the three typical recording situations. The first recording situation is a studio. Recording

studios are characterized by a very short reverberation time because of the large amount of absorption employed. Nevertheless, sound will still reach the microphone after only a single reflection as the surfaces are not perfectly absorbing. Because studios are usually not large, these initial reflections will arrive shortly after the direct sound. Later echoes are involved in more reflections, so are severely attenuated. For a recording studio the rationale is fairly clearcut.

Even when the recording is made in a more echoic environment, the same conclusion is reached. Typically, recordings are made with the microphones placed fairly close to the performers. There are three reasons why this is done. Placing the microphones close to the performers increases the signal level relative to the ineluctable noises due to air conditioning, traffic outside, planes overhead, and so on. The problem of background noise is particularly acute at live performances. Close miking also increases the signal level relative to the reverberation of the hall. Most recording engineers take this approach because of the economic pressure to expedite the setting up. Although it is possible to position microphones to provide an appropriate balance between the direct sound and the reverberance, doing so is usually time-consuming. Erring on the side of overreverberance would be fatal, whereas a too dry recording can be fixed by mixing in reverberation to achieve the desired effect. Another point is that without an audience a concert hall will be more live, so that the level of reverberance that would be recorded would be unrealistic anyway. The third reason for close miking is not as widely understood although it is actually the most important: close miking increases the signal level relative to the ambience. When the ambience of a hall is recorded stereophonically, most of the information that made it possible for us to resolve the direction of the reflections when we were present at the performance is lost. The resulting admixing makes the recording seem overly reverberant. By placing the microphones close to the performers, only the direct sound and the early ambience described before will be received. The reflections with longer delays produced by more remote reflecting surfaces will be greatly attenuated relative to this sound due to the greater length of their paths.

It is also possible, though more difficult, to make acceptable recordings with the microphones positioned in the audience seating area. The greater difficulty is reflected in the argument's being less clearcut. In this case, directional microphones must be used to exclude some of the natural ambience. The directionality of the microphones will favor echoes arriving

from the same direction as the direct sound. Simple geometry indicates that from this greater distance, these reflections will reach the listener at about the same time as the direct sound. The early reflections that arrive after a significant delay involve reflection off the side surfaces, and so are attenuated by the directional gain of the microphone. After an even greater delay, more sound will fall within the beam of the microphones, but the amplitude of these echoes will be attenuated by the distance travelled. So, in all three cases we can see how standard recording practice results in the recorded ambience being restricted to short times.

A related concern is what effect the transversal filter will have on the recorded reverberance. As noted before, reverberance is a noise-like signal, so that one of its most important properties is incoherence. A degree of incoherence exists between the two reverberance signals available in stereo, but their effectiveness is limited because they are emitted only from speakers in front of the listener. Adding the two stereo channels together and subjecting the sum to the complex filtering embodied by the simulator creates a third reverberance signal that is partially uncorrelated with the original signals. Emitting this new reverberance from the secondary channels increases the incoherence of the reverberance in the listening room, extracting greater benefit from the recorded signal.

The most significant problem with stereo recordings is that creating a convincing sense of distance from instruments that have been close-miked is sometimes difficult. When the direct sound provides cues that the source is nearby and the ambience corresponds to a source that is far, the cues conflict. This conflict often resolves in favor of the direct sound producing a simulation that is distorted—but not in a way that is grating. Despite this caveat, the auditorium simulator is remarkably effective with the vast majority of recordings.

5.4   PRACTICAL EXPERIENCE AND ADDITIONAL COMMENTS

Recorded performances auditioned with the simulator sound very life-like and natural. In contrast, stereo is flat and two-dimensional, and the sound is much more likely to be perceived as emanating from speakers. Although the simulator makes it possible to hear differences corresponding to subtle changes in the concert hall geometry, in practice it seems to be adequate to provide a small set of selections to cover the range of musical styles and personal tastes.

Selection of the desired acoustic environment is achieved simply by depressing the appropriate button. All the required adjustments are stored internally in the form of the impulse response so that users can deal with the simulator at a level that is easy to understand. This simplified control circumvents the criticism of previous ambience synthesizers for having controls that were unduly recondite. Another advantage of providing a limited selection of models is that listeners can hear a difference between different selections with almost any kind of music. This minimizes the confusion and frustration that can arise from controls that do not always produce an identifiable effect.

In the course of testing the simulator in different listening rooms, the setting up was invariably quick and easy. The only operations required after making the electrical connections are balancing the two ambience speakers to each other and balancing them with the stereo speakers. Because the adjustments are not critical, they can be completed with considerable celerity. The flexibility in positioning and orientation allows the secondary speakers to be as inconspicuous as possible.

It should be emphasized that unlike most previous units of this type, no recirculation of the sound is employed. Recirculation has been considered necessary in order to provide the gradual decay of sound characteristic of reverberance. As we have seen, generating reverberance duplicates information already present in recorded material. The extent to which previous units were effective is probably attributable in large measure to the accuracy with which they fortuitously created an appropriate early reflection pattern. Recirculation is completely unlike actual physical processes in concert halls, and it is very difficult to introduce without creating temporal patterns that produce a "springlike" quality. As a result, these devices often suffered from a cloying artificiality. Practical experience has shown that on units that provided the option, users often listened with the recirculation defeated. In this mode of operation only a few reflections would have been generated, with timing and amplitudes arrived at *ad hoc*. Some units that did not have provisions for recirculation provided only a single delay. With its more powerful hardware and careful utilization of its capabilities, this auditorium simulator accounts for as many as 100 early reflections. Perhaps more important than the greater accuracy is the greater naturalness. One can listen for hours without any fatigue.

5.5   CONCLUSION

Simulating auditorium acoustics in a domestic listening environment is a less demanding application that permits profound simplification. As before, the simulation is based on a detailed modeling of the acoustics of various performing spaces, but the directional dependence is less accurately conveyed. An inexpensive simulator convolves an audio signal with a selected impulse response in real time. Several impulse responses are stored to provide listeners a choice of environments to suit the musical style of their selection and their own personal taste. Only two additional speakers are required to convey the lateral early reflections, and their positioning is uncritical. The medial reflections do not contribute to the desirable spatial impression, so no resources are wasted presenting these sounds. Also, no effort is made to synthesize reverberance because this information can be and typically is recorded. The ambience created by this simulator is not merely an interesting effect, it is an important step toward greater realism.

CHAPTER 6

REFINING THE IMAGE MODEL

6.0 INTRODUCTION

Developing a concert hall model is an adaptive process. The model specified in the synthesis phase is evaluated in the analysis phase, and the results of this evaluation motivate improvements to the model. Successive iterations progressively refine the model. Ideally the analysis would indicate exactly what modifications to the model are necessary. That is usually the case when objective measures can be applied, but subjective measures rarely give more than a hint. Therefore, refining can be desultory, relying heavily on intuition and insight.

To approach the problem systematically, a researcher must begin by evaluating his assumptions. The cornerstone of this thesis is the image model. When an omnidirectional point source is near an infinite, smooth, rigid plane, the normal component of the particle velocity at the boundary must be nil, i.e.,

$$\bar{v} \cdot \hat{n} = 0. \tag{6.1}$$

In this case, an image source satisfies the boundary condition exactly allowing the boundary-value problem to be replaced by an equivalent problem involving two sources and no boundary.

There are many deviations from this simple ideal in real concert halls, but they fall into three broad categories: the characteristics of air, of the source, and of the boundaries. Refining the image model would logically begin by correcting these obvious, identifiable approximations. In this chapter, we will consider the nature of the errors that arise from ignoring these characteristics and suggest ways of enhancing the image model.

The unifying thread of this chapter is that additional effects can be modeled with filters. In the basic image model, the amplitude of the sound produced by a particular VS is weighted to take account of the attenuation it encounters along the path. The weighting is uniform with frequency. In effect, the sound for the ith VS is passed through a filter whose impulse response is $\gamma(t) = g_i\delta(t)$ (see Eq. 2.5). By allowing the impulse response to be a complex function of time, a frequency dependence can be introduced that takes account of many of the wave effects that have been neglected.

6.1   CHARACTERISTICS OF AIR

Air absorbs sound energy. Sometimes the distributed absorption is modeled by an exponential term [GIBBS-72], transforming Eq. 2.5 to

$$g_i = \frac{R_0}{R_i} \prod_{j \in S} \Gamma_j e^{-mR_i}, \tag{6.2}$$

where m is the attenuation constant of air in units of distance^{-1}. The value of the attenuation constant is a strong function of relative humidity and temperature. It is also a function of frequency, with high frequencies absorbed more than low [HARRIS-63][HARRIS-64]. Consequently, transmission through air introduces a progressive low-pass filtering with increasing distance. The effect of nonuniform sound absorption can be easily modeled by appropriately filtering the sound corresponding to each path. Accordingly, the transmittance for each path is augmented with a factor

$$\gamma_{air}(t; R), \tag{6.3}$$

where R is the length of the path. Moorer has proposed a simple first-order filter for modeling air absorption [MOORER-79]. How precisely the frequency dependence needs to be modeled is not

known. The fluctuations in the frequency response arising from interference between multiply-delayed sounds are very strong. If they overwhelm the gentler characteristics attributable to air absorption, the filters could be omitted. A less drastic simplification would be to use a single filter to model approximately the filtering for all the reflections, rather than customizing the filter characteristics for each one.

6.2   CHARACTERISTICS OF SOURCES

In approaching the problem of creating an audible simulation of a concert hall, we started with the premise that a concert hall could be considered to be a filter. By passing an audio signal through an electronic realization of the filter, we hoped to hear the same thing we would have heard were the sound emitted in a physical realization of the concert hall. But in Ch. 2 we encountered the first complication: sound reaches the listener from many directions. Because the subjective effect of a delayed sound varies with direction, the presentation must preserve directional cues. Accordingly, we must consider each unique VS direction to correspond to a different output of the filter (although the number of outputs can be reduced to two by accurately modeling binaural hearing).

In this section we observe that the complications are reciprocal: because the output of real sound sources varies with direction, we must also consider each unique VS direction to correspond to a different *input* to the filter. If the simulation is to match the sound the instrument or ensemble would have in a physical realization of the modeled concert hall, then the directional characteristics of the instrument must be taken into account. Determining from the image model the appropriate direction for each sound path is simple. Then in principle an accurate simulation could be obtained by recording the source in every direction required by the analysis. But just as a multitude of speakers was impractical for preserving directional cues in the presentation, so a multitude of recordings would be impractical for preserving the directional characteristics of the source.

The output of an instrument also varies with distance. But as long as we can assume that listeners will be in the far field, the variation with distance arises primarily from air absorption. We discussed the modeling of air absorption in the previous section. To be in the far field, one's

distance from the source must be large compared to its dimensions. There is an additional complication because we are allowing many sources to be playing at the same time. One must be farther from a large ensemble before near-field effects can be neglected. But as long as the source can be broken into small pieces, the total response can be considered the superposition of a number of responses so that far-field positioning is a valid assumption. Therefore, we can focus our attention on taking account of the variation with direction.

In comparisons, it is often possible to ignore some of these factors. A comparison is valid as long as both signals assume the same source characteristics. A source that comes closest to the implicit assumptions of the image model is an omnidirectional speaker, constructed by positioning drivers on the surface of a sphere. Alternatively, the effect of an isotropic radiator can be simulated with a directional speaker by averaging measurements in a large number of directions. However, the bandlimiting of speakers must be accounted for in the image model by an appropriate filter. Measuring the concert hall with such a source will provide a consistent basis for comparison, but some characteristics of the hall could be overlooked with such unnatural sources. Still, the most important characteristics of a hall can be determined with simple sources, so the enhancements to be discussed presently are required only for the most demanding applications.

6.2.1 Dependence on direction

To approximately model the radiating characteristics of real sources, we assume listeners are distant enough that the finite extent of the source can be ignored, making it possible to consider the source a point with nonisotropic intensity. A single recording preserves the sound of the instrument in a single direction, but a directional gain rule makes it possible to adjust the recording to obtain the sound of the instrument in any other direction. We characterize the instrument once in the laboratory by measuring its directional gain relative to the direction of the recording. Thus, the formula for the transmittance must be augmented with a term dependent on direction.

The procedure is complicated by the variation in directional gain with the note being played. Instruments are most directional at high-frequencies. This effect can be modeled by varying the gain for a particular direction with frequency. Thus, the directional impulse

response is no longer simply a weighted impulse, but a complicated function of time with the appropriate spectrum. To take account of the dependence on direction, the impulse response must also vary with direction, so that the transmittance is now computed from a formula augmented with the term $\gamma(t; \theta, \phi)$. Creating this directional gain rule would be very laborious because of the extensive measurements required. Furthermore, the model is still not perfect because it assumes a harmonic radiates in the same pattern as a fundamental with the same frequency. However, playing different notes subtly changes an instrument—by covering a different set of holes, for example—altering the radiation pattern. Another complication is that the radiating characteristics of instruments change with loudness. Because it is a nonlinear effect, this characteristic would be difficult to accomodate within the framework of a linear system. Nevertheless, using the directional gain would capture the essential characteristics of an instrument to a degree far surpassing any other technique, and would probably be accurate enough.

Reducing the measurements to a closed-form expression might be difficult. At the expense of greater memory requirements, an acceptable alternative would be to leave the data in the form of a *set* of impulse responses, just as is done in the dual problem (Sec. 4.3). In computing the impulse response of the concert hall, the direction of a VS would be quantized and used to select the appropriate impulse response from this set. Presumably, minute details of the dependence on direction can be ignored. Some smoothing would be appropriate in any case because of the expected variations from one instrument to another. Furthermore, precisely controlling the position of an instrument in any practical situation would be impossible, and besides, musicians usually move when they play. The effect of this smoothing would be a spatial bandlimiting, so the sampling theorem indicates that no information would be lost by sampling the dependence. If necessary, values at intermediate directions could be recovered by interpolation.

Another point to consider is the significance of phase. Comparing the phase at different frequencies is impossible because the instrument can sound only one note at a time. Therefore, only the magnitude frequency response for each direction is easily found. But assuming an instrument can sound only one note at a time ipso facto obviates the question of how sounds at different frequencies combine. The only important phase effects have to do with the relative outputs in different directions for a *given* note because they determine how the reflections

combine at the listener's position.

The most convenient measurement procedure would move a single microphone to different positions on the imaginary sphere surrounding the instrument, making it impossible to compare the phase in different directions. Phase is meaningful only in a comparison, so at least two microphones are required. Indeed, the intensity measurements can be made with a single microphone only by assuming the instrumentalist can maintain a constant amplitude, probably an unrealistic assumption. While a two-microphone procedure suffices, accurately accounting for the phase is unlikely to significantly affect the simulation. At higher frequencies, a given phase delay corresponds to a small change in the length of the sound path. As long delays are an inherent part of the simulation, small changes are insignificant. At low frequencies some instruments are omnidirectional, so we can assume a priori the same phase for each direction. Therefore, carefully accounting for phase would be a subtle improvement at most; the salient feature of the directional variation has to do with the changes in intensity with direction.

6.2.2   Effect of finite extent

Modeling performing groups as a point would seem to be a poor assumption, particularly in the case of a symphony orchestra. At least one would want to consider an ensemble to be a collection of point sources—a possibility if recordings are available of individual instruments. The image model could compute the auditorium response for each performer given his position. Convolving these responses with the sound of the corresponding instrument would produce the sound of the instrument at the listener's position. Then superposing the sounds for individual instruments would produce the sound of the entire ensemble. The result would be more accurate because it would take account of the differences in the response of the hall from different parts of the stage.

This approach would probably provide sufficient accuracy for even the most demanding applications, even though individual instruments are not point sources, particularly large instruments such as a piano. Bass violins and cellos have even larger effective radiating surfaces when they are acoustically coupled to the stage. At sufficiently large distances, the acoustic field produced by sources with finite extent is equivalent to the field produced by a nonisotropic point source [PIERCE-81]. The approximation is true at any frequency for which the listener

is a sufficient number of wavelengths from the source. For a full orchestra, this would be a poor approximation, but for individual performers the approximation would be valid except in locations very near the stage. (But there the characteristics of the hall are dominated by the direct sound.) Sections of an orchestra might be small enough in many cases.

When the approximation is not satisfactory, the only alternative is to measure the response not only as a function of direction and frequency, but also of distance. Such a scheme would be almost unimaginably cumbersome. The only practical solution is to break up the ensemble into small enough components that the point model is valid. If other restrictions make it impossible to satisfy this condition, the resulting distortion must be considered a fundamental limitation of the model.

6.2.3   Experimental paradigm

To accurately model the effects of source directivity, we propose the following experimental paradigm:

(1) Make all measurements under anechoic conditions.

(2) Determine the directional impulse response for an individual instrument by measuring the acoustic output at a large number of points on the surface of a sphere surrounding the performer. The surface of the sphere should be in the far field of the instrument. At each position, the acoustic output should be measured for a set of notes covering the range of the instrument. As we are trying to extract the amplitude versus frequency, the fundamental should be isolated by filtering. Note that special measures are required to determine the directional characteristics of the harmonics of the highest note. The measurements for each direction must be reduced to the impulse response for that direction. Then, if possible the full set of measurements are assembled into a formula for the impulse response as a function of direction.

(3) Repeat the process for other instruments.

(4) Assemble the sound for the entire ensemble from the responses for the constituent instruments.

(5) When computing the contribution of a particular VS to the impulse response, determine the direction of the listener and apply the directional gain rule to find the impulse response for

that direction.

(6) Use the distance of the VS from the listener to control another filter that models nonuniform sound absorption in the air.

In many cases, an important simplification that would probably be acceptable is to measure entire sections of the orchestra in steps 2 and 3 rather than individual instruments.

6.2.4 Stereo recordings

There is a strong temptation to use stereophonic recordings as the source material because they are readily available, but it should be eminently clear at this point that conventional recording methods are woefully inadequate for all but the crudest simulations. Stereo recordings have two prominent limitations. First, commercial recordings are always made in echoic environments. Although the close- or directional-miking techniques that are often used exclude much of the reverberation, the final product is usually supplemented with artificial reverberation. As discussed in the previous chapter, this recorded reverberance can be utilized constructively in crude simulations. But in more accurate simulations, the recording should provide only the direct sound, with all the reverberant sound created in the simulation.

The other important limitation is the confusion of directional cues. Stereo recordings are often made by blending the outputs of multiple microphones in a way that attempts to create a sense of the position of each instrument by controlling the balance of the sound fed to the two channels. In a crude simulation, one can assume the source is a point insofar as the reflections are concerned, and utilize these directional cues in the direct sound (Ch. 5). But the sound of an instrument changes in other ways in a concert hall: the response of the hall from different parts of the stage is different. The directional cues provided by stereo are insufficient for modeling these subtler but potentially important effects.

There are two other problems that arise with stereo recordings. The balance is often adjusted depending on which group of players has the melody, and soloists might be made more prominent. There is also the difficulty of creating a sense of distance from a source that has been close-miked to obtain the desired timbre. All these limitations can distort simulations in very objectionable ways.

6.3 CHARACTERISTICS OF BOUNDARIES

6.3.1 Local reaction

The basic image model assumes the boundary is rigid, i.e., the normal velocity is zero regardless of the incident pressure. The ratio of pressure and velocity at the surface is defined as the specific acoustic impedance:

$$\varsigma = \left(\frac{p}{v_n}\right). \tag{6.4}$$

Thus, a rigid surface has an infinite impedance. Actual boundaries have a finite impedance that is generally a function of frequency and angle of incidence.

In the image model, the acoustical properties of a surface are characterized in terms of the reflection coefficient. The impedance and the reflection coefficient are related by the well-known expression

$$\gamma = \frac{\varsigma\cos\theta - 1}{\varsigma\cos\theta + 1}, \tag{6.5}$$

where γ is the reflection coefficient and θ is the angle of incidence. Thus, the reflection coefficient is also a function of frequency and angle of incidence.

The image model usually assumes the reflection coefficient is independent of the direction of incidence. Although this assumption is not true in general, it is sometimes justifiable for diffuse reverberance. In that case, sound is equally likely to reach the surface from any direction, so that the average behavior can be studied by averaging the reflection coefficient over all directions. This rationale does not apply to the ambience because the reflections are relatively infrequent so statistical behavior does not apply.

Although it is usually a poor assumption that the reflection coefficient of the surface is independent of direction, in many cases the specific acoustic impedance is. Assuming an impedance independent of direction implies that the surface is "locally reacting." When a surface is not locally reacting, adjacent surface elements are coupled by bending stiffness, making it possible for it to sustain a traveling wave in a direction parallel to its surface. Non-locally reacting surfaces can produce important effects in concert halls. The stage is usually designed to be flexible so that it can serve as a sounding board for instruments such as bass violins and cellos, enhancing the output of an orchestra at low frequencies. Also, the audience seating area at orchestra level often has a wooden floor, further enhancing the transmission of

low frequencies to part of the audience. Allowing sound to propagate along a surface would be very difficult to model because of the distributed nature of the process. Fortunately local reaction is usually a reasonable assumption.

6.3.2 Finite impedance

Assuming the surface is locally reacting implies that the impedance is independent of direction. Eq. (6.5) shows that when ς is independent of direction, the reflection coefficient depends on $\cos\theta$. Computing $\cos\theta$ from the image model is straightforward. If the unit normal for the last surface that reflected the sound is $\hat{n}$, and the vector from the listener to the VS is $\overline{R}$, then

$$\cos\theta = \frac{\overline{R}\cdot\hat{n}}{|\overline{R}|}. \tag{6.6}$$

To find $\cos\theta$ for other reflections encountered along the same path, the process is repeated with virtual listeners of progressively higher order, VSs of progressively lower order, and the unit normal for the appropriate side in the same manner as the visibility testing (see Ch. 2).

Another characteristic of a finite impedance boundary is that the impedance generally varies with frequency. For example, real boundaries are pressure release at low frequencies— their impedance is -1—allowing low frequencies to be transmitted through the surface. Also, surfaces usually absorb high frequencies more readily than low. To combine the dependencies of the reflectance on direction and frequency, the impulse response is made a function of time and direction θ: $\gamma(t;\theta)$. The effect of multiple reflections is obtained by convolving the responses for each one.

When energy is absorbed at the boundary, the image model no longer satisfies the boundary condition Eq. (6.1). Thus a diffraction wave is required to satisfy the boundary condition. These more subtle effects of finite impedance boundaries are discussed in several recent papers [WENZEL-74][THOMASSON-76][DONATO-76].

6.3.3 Isolated finite planes

The effectiveness of an isolated plane in reflecting an incident sound wave depends on its dimensions compared to the wavelength of the sound. Less sound is reflected when the wavelength is much longer than the dimensions of the plane. Therefore, any finite plane should have a reflectivity with a high-pass characteristic. A recent paper [SAKURAI-81] modeled this effect by numerically integrating the Kirchhoff-Helmholtz integral theorem under the

assumption that the velocity potential behind the reflector is zero. The solution models the effects of a finite plane in terms of a "geometrical wave" and a "boundary wave," where the former corresponds to the specular component, and the latter models the effect of the finite extent. The solution indicates that when the VS is visible, there is an additional reflection after the geometrical wave when the wave front reaches an edge causing a boundary wave. The boundary wave has the effect of canceling sound energy at low frequencies, so the impulse response does indeed show a high-pass characteristic. The delay of the boundary wave depends on the excess length of the path from the VS to the listener by way of the boundary. As the boundary becomes larger, the boundary wave arrives later—moving the cutoff frequency lower—and weaker. Thus, the solution for the finite plane corresponds with the infinite plane case in the limit as the dimensions increase without bound. Similarly, in the limit of vanishingly small wavelength, the boundary wave vanishes leaving only the geometrical wave. Computed impulse responses correspond amazingly well to measured impulse responses for comparable cases.

Sakurai also deals with the situation of two nonintersecting finite planes [SAKURAI-82]. He argues that the numerical integration of the Kirchhoff-Helmholtz theorem effectively treats each element of the boundary as a point source. Thus, the effect of two nonintersecting planes can be calculated by considering each boundary element to be a secondary sound source. As such, each element is responsible for a geometrical and boundary wave reflection from the other surface. Likewise, the reflection from the second surface of the first geometrical wave and each component of the first boundary wave can be accounted for by considering the second geometrical and boundary waves for each component. Clearly, the computation grows rapidly as higher order reflections are considered, but the boundary wave of the second reflection corresponding to the boundary wave of the first reflection can be neglected in most cases because its amplitude is small.

6.3.4 Intersecting planes

As we noted in the introduction, the image model produces an exact solution for a rigid, infinite plane. When two semi-infinite, rigid planes intersect with an angle θ, the images will not always satisfy the boundary conditions at both surfaces simultaneously. To illustrate, Fig. 6.1 shows for two dimensions the image positions that result when the two sides meet at an angle of 120°. We can see that vs_a is required to satisfy the boundary condition for side 1, but the boundary condition for side 2 is still not satisfied. To satisfy the boundary condition for

SOURCE 2   SOURCE 1

SIDE 2   SIDE 1

a'•   • a

b •   • c
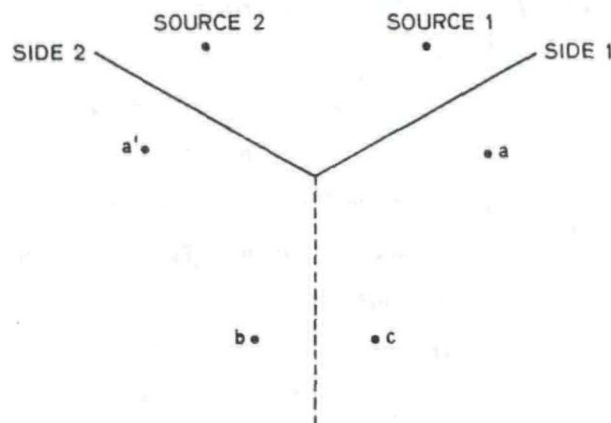
Fig. 6.1   Satisfying the boundary condition requires symmetry around a side. When two planes intersect with an angle of $120°$, both boundary conditions can be satisfied only when another source is positioned symmetrically as shown.

side 2, we must introduce the other image sources vs_b and vs_c, but then the balance at side 1 is destroyed. To satisfy the boundary conditions at both boundaries simultaneously, there must be perfect symmetry around each side. When the angle of intersection is $120°$, symmetry can be created only by positioning a second source as shown, or by restricting the one source to the bisector.

When the planes meet at a right angle, three virtual sources are required, as shown in

SIDE 2

a   S
•   •

VIRTUAL
SIDE 1 - - - - - - - - - -   SIDE 1

•   •
c   b
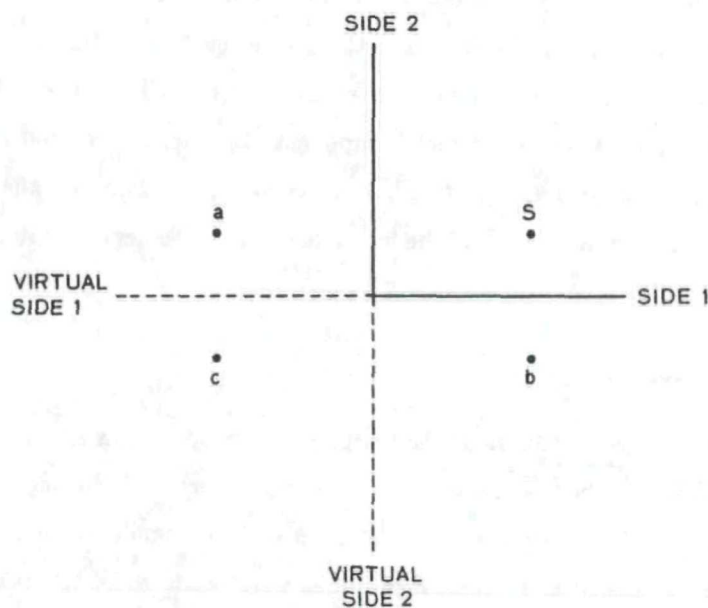
VIRTUAL
SIDE 2

Fig. 6.2   The familiar case of two surfaces intersecting at right angles shows how the boundary conditions will be satisfied by the VSs shown regardless of the position of the source. Only when a surface has a virtual extension that is colinear will this be possible.

Fig. 6.2. The perfect symmetry around both sides assures that both boundary conditions are satisfied regardless of the position of the source. The question arises for what other geometries the image model provides an exact solution. Clearly, the only way perfect symmetry can exist regardless of the position of the source is when each side has a virtual extension that is colinear. In other words, the half-plane must be divided into an integral number of sectors, or $\theta = \pi/r$, where r is some integer greater than 1.

Having found the condition for a two-dimensional wedge, we next determine for which polygons the image method provides an exact solution. First, we need to determine the total internal angle of a polygon with n sides. Consider Fig. 6.3. From any internal point, the polygon can be divided into n triangles. Each triangle has a total of π radians, so the triangles have a total of $n\pi$ radians. But 2π radians are due to the rotation around the central point, so the polygon has a total internal angle of $n\pi - 2\pi$. Vertex k must have an angle of π/r_k, so

$$\sum_{k=1}^{n} \frac{\pi}{r_k} = n\pi - 2\pi. \tag{6.7}$$

Canceling the π from both sides we have the desired condition:

$$\sum_{k=1}^{n} \frac{1}{r_k} = \frac{1}{r_1} + \frac{1}{r_2} + \cdots + \frac{1}{r_n} = n - 2. \tag{6.8}$$

One obvious solution is a rectangle, for which all the r_ks equal 2 and $n = 4$. There can be no n larger because incrementing n increases the right side by 1 whereas the left side increases no
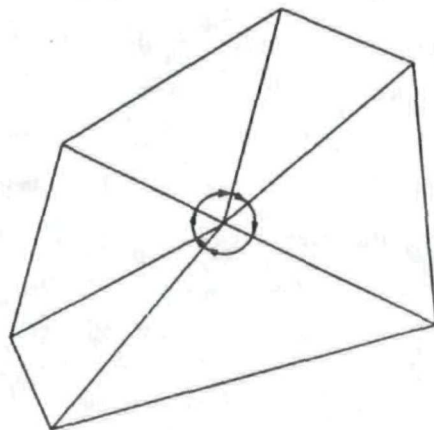
Fig. 6.3 To determine the total internal angle of a polygon, divide it into triangles by drawing lines from any internal point to each vertex. Each triangle contributes π radians less the amount at the center.

more than 1/2. The only other possible value for n is 3, and there are three solutions in this case:

$$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$$
$$\frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1 \qquad (6.9)$$
$$\frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1.$$

Only for these three triangles and a rectangle will the image model provide a complete solution.

In extending these results to three dimensions, there are several possibilities. The most straightforward is a half-infinite prism, where one of the permissible polygons is extended normally in one direction. Keller [KELLER-53] claims that the only permissible closed polyhedra are tetrahedra, the finite right triangular prism, and a rectangular parallelepiped. In fact, Allen and Berkley [ALLEN-79] have mathematically demonstrated the correspondence between the image model approach and the direct wave solution for the latter case. In all other cases, the image solution does not satisfy the boundary conditions.

What is wrong with the image-model solution for other polyhedra? The position of a VS still accurately indicates the direction from which most of the sound reaches the listener. That the boundary condition still is not satisfied indicates there must be another source of sound, and the additional sound supplements the contributions of the VSs so that the boundary condition *is* satisfied. The additional sound source is diffraction.

Whenever the VSs do not satisfy the boundary conditions, sharp edges can be found in the sound field they produce. For example, Fig. 6.4 shows the case of two sides intersecting at a slightly obtuse angle. Each second-order VS produces a wave propagating only in the region where it is visible. Nature does not allow a wave to have a sharp edge, so diffraction patterns exist near the boundaries of the regions. Restated heuristically, when the listener moves around, the image model has VSs change abruptly between visibility and invisibility, whereas the change actually happens smoothly. The extent of the diffraction patterns decreases with the wavelength of the sound, so geometrical acoustics provides an exact solution in the limit of vanishing wavelength. As long as one is far from an edge, the image solution is accurate.

6.3.5 Diffuse reflection

The image model assumes reflections are specular, a valid assumption when the surface irregularities are small compared to the wavelength of the sound. Unfortunately, this condition
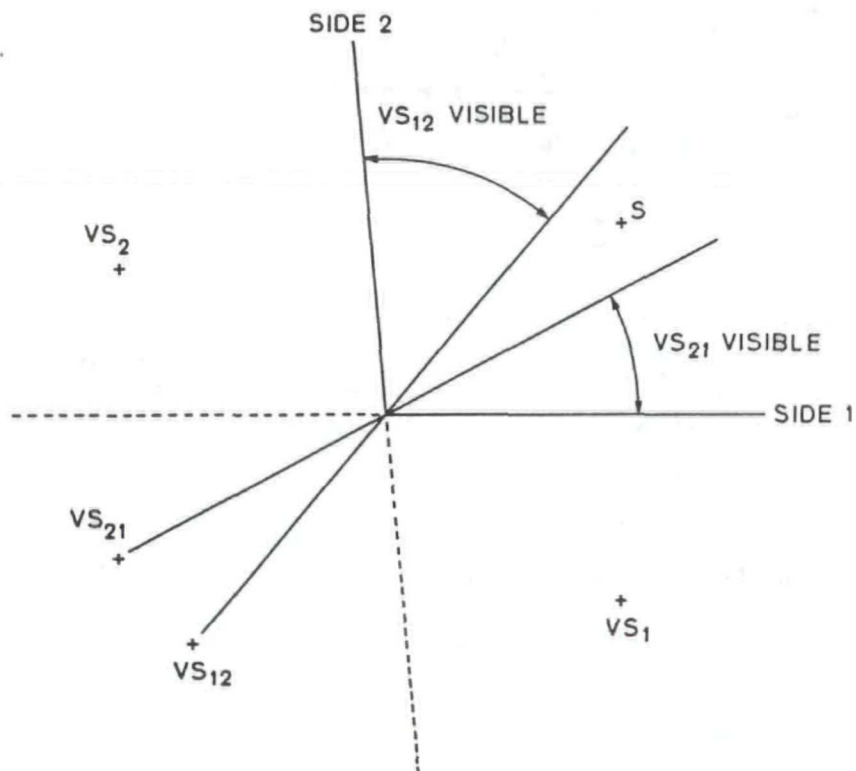
Fig. 6.4 When two sides intersect at an angle that is not π/r, edges can be found in the sound field. These edges must have diffraction patterns associated with them, so in their vicinity the image model solution is incomplete.

does not always apply in a concert hall. In fact, architectural acousticians often go to great pains to assure that reflections are not specular. Diffuse reflection arises from a variety of architectural features. Some are explicit efforts, such as the multifaceted ceiling designed by Harris for Avery Fisher Hall [BLIVEN-76] or Schroeder's maximal diffusors [SCHROEDER-79a]. Other diffusors seem almost to have been introduced inadvertently, such as the statues along the walls of Symphony Hall. A comprehensive mathematical model for a concert hall must be able to deal with diffuse reflections because they are an important physical process.

Diffuse reflection has sometimes been modeled simply as an additional absorption effect. Part of the incident sound is reflected away from the specular path, so clearly the intensity of the specular component must be reduced relative to specular reflections. But this model is incomplete because it ignores sound reaching the listener along nonspecular paths (Fig. 6.5). A complete characterization must consider not only the effect on the specular component, but also the contribution of the nonspecular sounds.
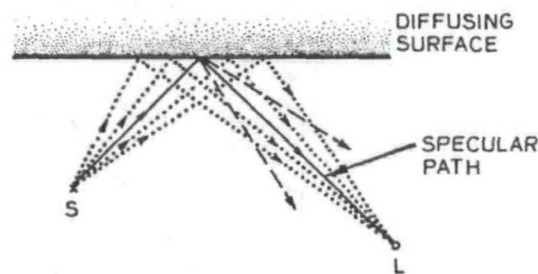
Fig. 6.5   The intensity of the specular component of a diffuse reflection is lower than it would be were the surface specular because part of the incident sound is reflected in other directions. But a signal processing model must also consider that sound can reach the listener along nonspecular paths.

In some cases, the diffusing effect can be modeled explicitly. For example, Avery Fisher Hall has diffusing facets that are all planar. Specifying each facet explicitly would accurately model their diffusing effect (although the effects of diffraction would still be neglected), but only at the expense of tremendous additional computation.

As usual, the impulse response embodies the desired signal processing model. Considering the surface to be flat expedites the computation of the VSs, and then the effect of diffuse reflection is superimposed. The general characteristics of the impulse response are easy to deduce. First, the diffused sound travels a longer path to the listener, so it reaches the listener later than the specular component. Second, its amplitude is reduced relative to the specular sound because (1) the extra distance traveled reduces the amplitude according to the spherical spreading law, and (2) the efficiency of a diffusor is lower for directions far from the specular. The later the sound the lower its intensity, so the impulse response of a diffusing surface will have a generally decaying characteristic.

We can see that the effect of a diffusing surface can be accounted for in the image model by introducing the appropriate filtering, where the characteristics of the filter are related to the acoustical properties of the surface. Similarly, the effect of multiple diffuse reflections could be modeled as a cascade of filters. The image model identifies the specular paths, and a filter is associated with each path according to the number of reflections and the characteristics of the surfaces encountered. The desired refinement of the image model amounts to augmenting the calculation of the transmittance for each path with an additional factor.

In general, the degree of diffusion depends on the frequency of the signal. Diffusing lower frequencies would require large surface irregularities so there is typically less diffusion at low

frequencies. The frequency response of a specular surface is flat, so we should expect diffusion filters to be flat at low frequencies.

As usual, there are complications. First of all, the appropriate filter has the unusual feature of a spatial dependence. That is, the diffused sound reaches the listener not only at different times than the specular sound, but also from different directions. A filtering model effectively reduces the process to one dimension. How important the spread of diffused sound is subjectively is difficult to guess. It is unlikely our acuity to the direction of diffused sound is as great as for isolated sounds because our perception is probably dominated by the specular component, which is strongest. However, when a surface is highly diffusing, or when many successive reflections amplify the spread, one could easily imagine that the spatial dispersion would eventually be large enough to be perceptible. Further, having the sound reach the listener from many directions is one way of reducing the interaural crosscorrelation which is known to contribute to a sense of spaciousness, so spatial diversity could be important. Taking account of the directional dependence would be possible—though laborious—by subdividing the impulse response into segments and conveying the direction of each one, using the binaural simulation described in Ch. 4. However, such complications destroy the elegant simplicity of the filter model, so should be attempted only after conclusively proving that the effect is subjectively significant.

Another difficulty with the filter model is that it treats the surface uniformly, that is it assumes the effect of the surface is independent of where the specular component strikes. Thus, errors will arise when the sound from a VS passes near a corner. There are two cases to consider (Fig. 6.6). When a VS is just visible, some of the delayed sound arising from diffusion corresponds to paths that are actually intercepted by the adjacent surface. This error can be corrected by modifying the filter characteristic according the proximity of the intersection point to an edge, but only at the expense of considerable additional computation.

Conversely, when the VS is just invisible, no energy will arrive from its direction even when some of the diffused sound should. Diffusion is accounted for only by associating its sound with a specular path. When the specular path is discarded, the associated diffused energy is lost as well. Dealing with this error would require detecting in the visibility test when a VS is just invisible and adjusting the filter appropriately, again sacrificing the simplicity of the model.

Obtaining the appropriate impulse response for the diffusing surface is a profound difficulty.
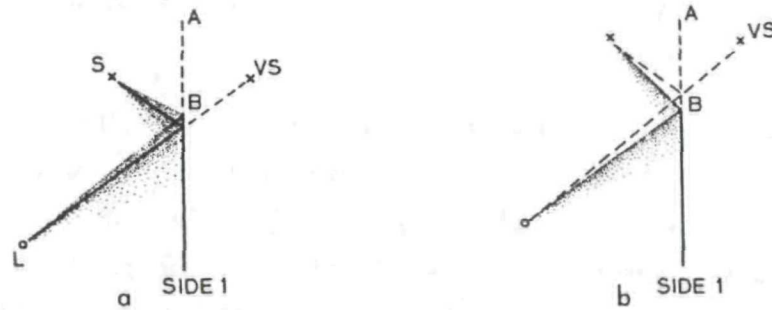
Fig. 6.6 Two types of error can arise when using the simple diffusion model. In the first case, part of the diffused sound is excised when the side extents only to B. The filter model treats the surface uniformly, so this effect is missed. In the other case, some diffused sound reaches the listener from a VS that is invisible. Because the diffusion filters are associated with specular paths, this sound is completely omitted.

One could measure a sample of the diffusing surface in an anechoic environment, but such a measurement would not isolate the effect of diffusing reflection, if that is a consideration. Techniques for mathematically modeling the effects of diffusing surfaces are an area of active research [WATSON-83a][WATSON-83b] [TOLSTOY-79][BIOT-68][BURKE-66].

6.4 CONCLUSION

The basic image model presented in Ch. 2 is general insofar as geometry is concerned, but it ignores many wave effects that can be important in concert hall modeling. In this chapter we considered several having to do with the absorbing properties of air, the radiating characteristics of sources, and the reflecting properties of boundaries. It appears that additional physical processes can be modeled by filtering the sound for each particular path. The cumulative effect is obtained by convolving the impulse responses for each process. Compromises minimize the impact on the computational efficiency, although it is still important that only those effects shown to be subjectively significant be considered.

CHAPTER 7

CONCLUSIONS

7.0 SUMMARY

In this thesis we have discussed a method for developing a model of a concert hall. Where previous work has sometimes been desultory, we have tried to be systematic. Thus we considered the problem to comprise two phases, synthesis and analysis. The synthesis phase—defining a model—developed an extension to the image model that allows it to deal with complex geometries closer to those actually encountered in practice. This model is useful for analytical studies as well as for creating audible simulations.

One application of the extended image model to analytical studies is establishing the connection between concert hall geometry and subjective response. Using the extended image model, we were able to explain why rectangular concert halls sound better than fan-shaped ones. In the fan-shaped hall, the VSs are compressed toward the median plane, decreasing the degree of spatial impression. Furthermore, there is another shape, the reverse fan, that might be even better than the shoebox because it spreads the VSs farther to the sides. Although the reverse fan suffers from some practical limitations, it is possible to obtain the same acoustical

performance with a segmented fan in which the segments are positioned according to the fan shape but oriented according to the reverse fan. Allowing the segments to rotate would provide a much more meaningful control over the acoustical characteristics of the hall than techniques currently in vogue.

Analysis of this model began with measurements of existing halls to provide a standard for comparison. Conventionally, an impulse is used to measure the impulse response, but its limited energy makes it difficult to overcome ambient noise. Noise is more suitable: being a sustained signal, it contains more energy than an impulse. The impulse response is extracted from the measurement by crosscorrelating it with the original excitation. A noiselike excitation based on a binary maximal-length sequence allows this crosscorrelation to be performed efficiently. Because the excitation is binary, only additions are required. Dramatic simplifications result using the algorithm we derived that uses the fast Hadamard transform to minimize the number of additions. A straightforward crosscorrelation would require approximately n^2 multiplications. But this efficient algorithm requires only about $2.5 \log_2 n$ additions, which represents a reduction in the execution time of several orders of magnitude for most values of n.

Having measured and modeled the same hall, the next step was to compare the two impulse responses. Chapter 4 proposed a method of subjective comparison to avoid the difficulties of objectifying subjective response. Directional cues were preserved by basing the presentation on binaural hearing. The binaural impulse response of the existing hall is easily measured by placing microphones in the ears of a subject or a dummy head. But computing the binaural impulse response required an extensive set of measurements of the directional impulse response of a subject's ears. The position of a VS relative to the listener is used to select the appropriate binaural impulse response so that in the simulation the corresponding sound seems to reach the listener from the appropriate direction. This approach is very general and precise, and does not require making assumptions about the properties of hearing that may prove false. The precision of the binaural presentation makes the approach attractive in demanding research applications.

Domestic audio reproduction is a less critical application, allowing simplifications that minimize the expense without completely sacrificing accuracy. The most important simplification has to do with the method of presentation. Speakers are used to present the delayed sounds, but

their number is minimized by ignoring elevation and by positioning speakers so as to convey many sounds in approximately the same direction. Medial sounds are ignored because they do not contribute to spatial impression. Other simplifications were assuming symmetry and combining closely-spaced echoes. Together, these assumptions make it possible to deal with over 100 echoes while explicitly processing only 15. Using a logarithmic attenuator matches the properties of hearing and minimizes the expense of the digital signal processing circuitry. The results obtained are very natural and life-like. They give the reproduction a three-dimensional quality and make the listener feel that he is enveloped in sound—certainly a worthwhile enhancement of commercial audio reproduction.

The thesis concluded with a chapter suggesting refinements to account for physical processes currently ignored. The limitations fall into three categories, the characteristics of air, the source, and the boundaries. Although the presentation was cursory, we showed that it should be possible to model many of these effects by incorporating additional filters. However, the computational burden imposed by the refinements makes it crucial that only physical processes that can be shown to be subjectively important be incorporated.

7.1   THE NEXT STEPS

7.1.1   Refining

We have outlined a procedure for developing an accurate model of auditoriums involving successive synthesis and analysis. In this thesis we have made a first pass; the next step is to apply some of the ideas of Ch. 6 to synthesize a refined model, beginning the second iteration. Subsequent evaluations and refinements would progressively refine the model.

Measurements of the directional radiating characteristics of musical instruments would be an important first step. Although some information has been published [JANSSON-75], the treatment has been far too superficial for this application. Allowing the frequency response to be controlled independently for each path is another potentially important enhancement. However, one would want to ascertain that the improvements were audible because additional computation slows the processing.

There does not appear to be any important limit to the accuracy that can be achieved with computer modeling. We feel that these techniques will become increasingly important in the future in advancing the sciences of architectural acoustics and psychoacoustics. Within a short time, electronic simulation of auditorium acoustics could make fiascoes like Philharmonic Hall a thing of the past.

7.1.2   Electroacoustic enhancement of auditorium acoustics

Many concert halls, particularly those with a fan shape, have an inadequate early reflection pattern. In such a case, the best solution is often to reposition the side walls, a remedy that is rarely practicable. A different approach is to synthesize lateral reflections using electroacoustical techniques. Electronic signal processing makes it possible to break the reliance on architecture to delay the sounds produced by the performers. Emitting these electronically delayed sounds from appropriately positioned speakers will make it possible to synthesize the early reflection sequence that listeners find most pleasing. It might also be possible to eliminate undesired reflections by emitting an appropriately delayed signal with inverse phase so that it cancels the offending reflection. Although this proposal shares much of the instrumentation of conventional public address systems, it goes far beyond such systems in that the sounds injected into the room are carefully derived and introduced according to new psychoacoustic principles. Ideally the result of the modifications would be an early reflection pattern similar to concert halls with highly regarded acoustical properties.

This approach to electroacoustic modification has several other applications. In order to maximize the utility of a new hall, features have often been provided for adjusting the acoustics to accomodate different musical styles. One way this is done is by suspending sound absorbing material from the ceiling in a way that allows the amount of material exposed to be adjusted. Although it is possible, using these techniques, to adjust the reverberation time over a large range, the subjective effect is often quite subtle because there is little change in the early reflections. A shorter reverberation time is noticeable only when the music stops, whereas the early reflections, which are at a much higher amplitude, have an important subjective effect almost all the time. Modifying the early reflection pattern is much more meaningful and will make it possible for the first time to adjust the acoustics of halls in a subjectively significant way.

Another advantage of electroacoustic enhancement is that it would give architects greater freedom when the design assumes their application from the outset [FISCHER-83]. This freedom could be used to solve problems that seem to be insoluble given the constraints imposed by the acoustical considerations. For example, it might be possible to consider radically different geometries that accomodate many more listeners without compromising the degree of visual intimacy with the performers. Also, it would be possible to separate the design of the acoustics in the listeners' area and the stage. Finally, electronic reinforcement would increase the loudness of large halls to a desirable range. Electroacoustical synthesis separates the problems of architecture from the problems of acoustics, making it possible to optimize each one.

Attempts to modify the acoustics of a concert hall must recognize the dichotomy between ambience and reverberance. Correction of the ambience requires care because details can influence our perception. Cruder techniques suffice for the reverberance because of its noise-like characteristic. Previous work has almost always focussed on modification of the reverberance [PARKIN-75]. As concert halls are often deficient in the early decay portion, efforts that disregard this region can be expected to have only limited success.

Modification of the acoustics must begin with suitable measurements. One straightforward approach is to use a directional microphone. Jaffe Acoustics used a parabolic dish to limit the view of their measurements to a 30° beam [JAFFE-81]. If their method is found to be inadequate, then beamforming might be the answer. In beamforming, the room is measured with a three-dimensional array of microphones. Combining their outputs with various delays creates the desired directional patterns [BROADHURST-80][ROCKMORE-unp]. A microphone array has the potential advantages of greater control over the directional pattern, more uniform response over frequency, and less disturbance of the sound field being measured. Ultimately it would be desirable for the measurement of the directional impulse response to incorporate established features of auditory perception, thereby simplifying the measurement and minimizing the amount of subsequent processing required.

These measurements of the directional impulse response will have four applications. First, it will clearly be necessary to measure the behavior of the hall to be modified to determine what modifications are necessary. The second application is in measuring the characteristics of concert halls considered to have superior acoustical properties so that these halls can be used

as a paradigm for the modifications. A third application is in psychoacoustic studies involving modification of the measured impulse response. By comparing simulations before and after modification, this technique makes it possible to correlate objective features of reverberation with subjective response using very realistic impulse responses. The fourth application is in preauditioning the proposed electroacoustic modifications. For example, one could evaluate a particular speaker array without placing any speakers in the concert hall by superimposing on the measured impulse response the appropriate binaural impulse responses for the proposed speaker positions.

Additional psychoacoustic research might be needed to clarify certain aspects of subjective preference. Different concert halls usually differ in many regards, so it is difficult to identify the features of the measurements that account for the preferability of the hall. A very important tool in these experiments will be our technique for generating sound fields with arbitrary directional characteristics. For example, it might be interesting to repeat some of Barron's experiments dealing with the directional dependence of spatial impression [BARRON-81] using complex reflection patterns that are closer to those actually encountered in concert halls. This will now be possible without requiring a large number of speakers.

By also applying sophisticated signal processing algorithms, it will be possible to quantify various aspects of auditory response that have been difficult to study with conventional instrumentation. For example, studies of the importance of spectral modifications in the delayed sounds will be greatly facilitated. Spectral modification will certainly be a part of the processing because it is a characteristic of real reverberation. But it is important to understand what constitutes desirable spectral modification and what would be undesirable. This investigation is also important for defining the capabilities required in the loudspeakers used to emit the synthesized sounds in the concert hall.

The second category of investigation concerns the implementation. In this category are such questions as how many microphones and speakers should be used and where they should be placed. Also, the characteristics of the speakers in terms of bandwidth, directionality, and power handling must be specified. Although the theory of the signal processing required is well known, the final application in the hall demands real-time operation. If the processing is complex, real-time operation could be challenging. A practical system might require some

special integrated circuits for the signal processing. Finally, a detail that must not be overlooked is designing the controls so that they are not too complex for technically unsophisticated people to operate.

The principal difficulties are deciding where to position the microphones and speakers, and perhaps also defining the objective for the modifications. Although measurements in good halls could serve as a paradigm, it is doubtful that a precise re-creation will always be possible. Then there would still be a question about the best achievable reflection pattern. The basic question about microphone placement is whether they should be positioned in the reverberant field or close to the performers. In principle, the reverberant sound can be obtained by electronically processing the unadulterated sound of the performers, but using the sound of the hall would have practical advantages. Speaker positioning is a problem because the speakers must serve all the listeners within earshot effectively. The sound reaching listeners close to a speaker must not be unrealistically loud, yet it must still be audible to those farther away. Also, because the subjective effect of delayed sounds varies with azimuth, the sound must reach listeners from the right direction to be constructive. Electroacoustic enhancement is the next logical step in the development of architectural acoustics.

APPENDIX 1

DENOISING

A1.0 DESCRIPTION

All audible tests were performed with the well-known and widely-used recording of dry orchestral music made by the BBC. Unfortunately, the recording suffers from excessive broadband noise. Noise causes confusion in comparisons because it is changed by the processing along with the signal. A new digital signal processing technique called "denoising" makes it possible to eliminate this broadband noise.

Denoising is a single-ended noise elimination technique. The familiar complementary techniques—Dolby and Dbx, for example—would not be applicable because the signal had not been appropriately encoded. Single-ended noise reduction is often performed by a mute circuit. A mute circuit measures the average amplitude of the signal and reduces the gain—abruptly to zero in most implementations—when it falls below a threshold. This low-level expansion increases the apparent output dynamic range. The familiar difficulties are that some low-level signals get trapped along with the noise, and that some noise modulation is usually audible. Reducing the gain smoothly for low-level signals would help hide the effect of the processing,

but muting is a crude technique which is rarely adequate.

Better results can be obtained by treating different portions of the spectrum separately. The Dynamic Noise Reduction (DNR) circuit marketed by National Semiconductor is a low-pass filter whose cutoff frequency changes depending on the high-frequency content of the signal, the cutoff moving to lower frequencies when the signal has little high-frequency energy. There is little audible filtering of signals rich in high frequencies, but the presence of the high-frequency signal masks the low-level noise in this case. Considering its simplicity, DNR is remarkably effective.

Even more noise reduction can be obtained by extending the filtering to bands between high energy regions. In effect we extend the variable lowpass filter to a variable bandpass, where the pass bands are dynamically varied according to the characteristics of the signal. It is still necessary to rely on masking in regions of the spectrum with high energy. Because the gain varies synchronously with the signal, these dynamic filtering techniques are actually a form of nonlinear distortion, but they have the subjective effect of reducing the apparent noise level.

Denoising operates in two phases. In the first phase the processor is instructed how to distinguish between noise and signal. A sample of noise must be available for this training. The sample is analyzed to produce an estimate of the noise power spectral density. It is assumed that the noise is stationary so that the same noise power spectrum can be applied in regions where signal is present.

The actual denoising is performed in the second phase using short-time spectral modification [CROCHIERE-83]. First, the signal is windowed and the resulting "frame" is transformed to the frequency domain. Then the magnitude of its spectrum is compared to the magnitude of the noise spectral estimate. Regions of the spectrum that fall below a threshold related to the noise power spectrum are reduced in magnitude with the phase preserved. After the short-time spectrum has been adjusted, it is returned to the time domain and spliced with previous results by another windowing operation. The operation can be compared to a large set of muting circuits each operating on a single band of frequencies.

The threshold for low-level expansion must be set higher than the noise spectral estimate

because approximately half the noise spectrum will exceed the estimate in any one frame. But even with any reasonable choice of threshold, it is possible for the spectrum in a narrow frequency band to poke above the threshold on occasion because of the random nature of noise. This problem is especially prominent in frequency ranges where the spectrum is crossing the threshold. It is also exacerbated by the nonstationarity of noise in actual tape recordings due to "self-biasing," where the signal itself provides additional bias, increasing the noise spectrum in regions of high signal energy. When neighboring frequency elements are below the threshold, the result of the denoising operation will be a narrow spike of noise, corresponding to a sinusoidal component in the time domain. Because the noise fluctuates, it is quite possible that in previous and subsequent frames the narrow frequency band was below the threshold. Thus, the sinusoidal components often last only as long as one frame—usually around 30ms. The short tones that result are called "breebles." They are often more objectionable than the original noise because of their distinctly tonal and rapidly changing character.
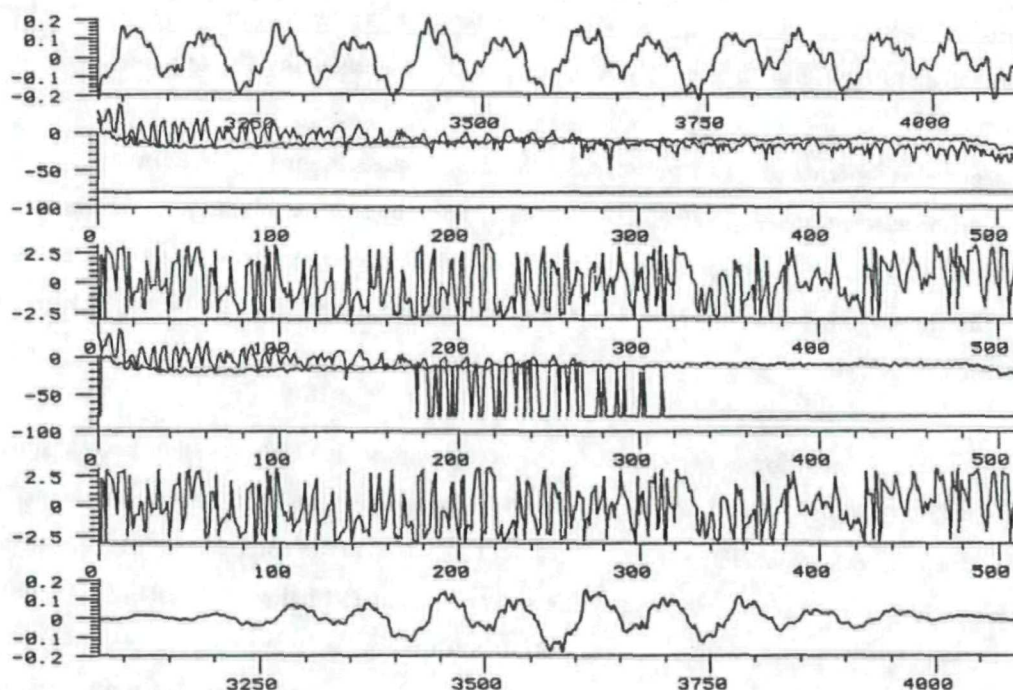
Fig. A1.1 This complicated display illustrates the action of the denoiser. The top trace is the input signal in the time domain. It is windowed and transformed to the frequency domain. The magnitude of the resulting transform is shown in the second trace along with the low-level expansion threshold and the expansion floor. Points below the threshold are slammed to the floor unless they are an isolated instance. The third trace is the phase of the spectrum. The next two plots show the resulting spectrum after the denoising operation is performed. The crossover region is where breebles are most likely to form. The phase is unchanged in the spectral manipulations, as seen in the next plot. Finally we have the resulting time waveform, to be spliced with previous results by a window and add operation.

There are several techniques for minimizing the audibility of breebles. First it should be stated that the problem is minor when the original audio signal had reasonable dynamic range to begin with, as is the case with the BBC tapes. The problem is objectionable only with signals that started with very poor dynamic range. The first remedy we tested had to do with the low-level expansion algorithm. The simplest choice is to reduce the magnitude of points below the threshold to the same small value. We call this the "slam" algorithm. A less abrupt algorithm "stretches" the spectrum proportionally to the amount by which the threshold exceeds it. For example, a logarithmic stretch reduces the magnitude by n dB for each dB below the threshold. Because this algorithm does not completely remove neighboring energy on either side of the noise spike, the spike is less tonal. Of course there is a penalty: less noise is eliminated.

Another technique we considered applies a higher level of intelligence. By storing several consecutive frames, it is possible to detect when a frequency bin is above the threshold for only a single frame. It is very unlikely for natural music to behave in this manner, so we arbitrarily suppress the spike under the assumption it must have been noise. This technique is remarkably effective, even with a buffer of only two frames. Obviously the same principle can be extended to the frequency dimension, at least at higher frequencies, as it is unlikely that a natural sound will have energy confined to frequency bands whose Q is higher than a certain amount. This refinement was not implemented as the first proved adequate.

APPENDIX 2

GENERATING COLORED NOISE

A2.0 INTRODUCTION

Ch. 3 described an efficient algorithm for measuring the impulse response of a linear system. Using noise as the excitation rather than an impulse allows more energy to be injected into the system, overcoming the background noise. The impulse response is extracted from the output by crosscorrelating it with the input. Two features of the algorithm account for its efficiency. First, the multiplications in the crosscorrelation computation are eliminated by using a binary sequence for the noise. But more significantly, generating the noise from a maximal length sequence minimizes the computations because then the crosscorrelation can be performed with the fast Hadamard transform.

Convolution is closely related to correlation, so an almost identical algorithm can be used to efficiently filter noise. It is often preferable to excite a system with a noise signal whose power spectral density is not flat. Pink noise, whose power spectral density has a $1/f$ dependence, is useful in acoustical measurements because the power is the same in each octave. Colored noise can also be used to provide a more constant dynamic range in measurements of systems whose

120

noise floor is not flat—for example, when preemphasis is involved or when $1/f$ noise is a factor.

Conventionally, noise with any desired power spectral density is obtained by passing white noise through a filter. The resulting power spectral density will be the squared magnitude of the filter transfer function. Clearly, any filter realization can be used to color the noise, but when long noise sequences are required, efficiency is a paramount concern. A maximal length sequence produces white noise, as desired. Because it is binary, efficiency considerations favor a FIR filter: the filtered output is the sum and difference of filter coefficients, whereas an IIR filter would require multiplications as well. As before, the number of additions and subtractions can be minimized using an algorithm based on the fast Hadamard transform. In this appendix we describe the efficient algorithm.

A2.1   FAST CONVOLUTION

The key to this efficient algorithm for coloring white noise is the relation between convolution and correlation. The crosscorrelation of two sequences $w(k)$ and $y(k)$ is defined by

$$\phi_{wy}(k) = \frac{1}{n} \sum_{j=0}^{n-1} w(j)y((j+k)), \tag{A2.1}$$

which, by a change of indices, can be rewritten as

$$\phi_{wy}(k) = \frac{1}{n} \sum_{j=0}^{n-1} w((j-k))y(j). \tag{A2.2}$$

As before, the notation $((j))$ indicates the residue of j modulo n. The crosscorrelation can also be expressed in matrix terms as

$$\Phi_{wy} = \frac{1}{n} \mathbf{W}_n Y, \tag{A2.3}$$

where Φ_{wy} and Y are vectors whose elements are the $\phi_{wy}(\cdot)$ and $y(\cdot)$ of Eq. (A2.2) and $\mathbf{W}_n$ is a right circulant matrix containing the shifted noise sequences. The convolution of $w(k)$ with the filter impulse response $f(k)$ is given by

$$y(k) = \sum_{j=0}^{n-1} w((k-j))f(j), \tag{A2.4}$$

which has the same form as Eq. (A2.2) except for the substitution of $w((-k))$ for $w((k))$ and the omission of the normalization factor. In matrix terms, the effect of negating the index is to transpose $\mathbf{W}_n$, so the convolution can be expressed as

$$Y = \mathbf{W}_n^T F. \tag{A2.5}$$

The terms of $\mathbf{W}_n$ are ± 1, so the matrix multiplications of Eq. (A2.3) and Eq. (A2.5) require only addition and subtraction, but the number of operations is proportional to n^2. The number can be minimized by substituting for $\mathbf{W}_n$ the sequence of matrix multiplications

$$\mathbf{W}_n = \mathbf{P}_2 \mathbf{S}_2 \mathbf{H}_{n+1} \mathbf{S}_1 \mathbf{P}_1. \tag{A2.6}$$

$\mathbf{S}_1$ and $\mathbf{S}_2$ suppress the first column and row of $\mathbf{H}_{n+1}$, and $\mathbf{P}_1$ and $\mathbf{P}_2$ permute the ones that remain. $\mathbf{H}_{n+1}$ is a Sylvester-type Hadamard matrix. Involving $\mathbf{H}_{n+1}$ in the matrix multiplications of Eq. (A2.3) is what accounts for the efficiency of the crosscorrelation because there is an efficient algorithm for multiplying by a Hadamard matrix known as the fast Hadamard transform (FHT). Substituting Eq. (A2.6) into Eq. (A2.3) produces

$$\Phi_{wy} = \frac{1}{n} \mathbf{P}_2(\mathbf{S}_2(\mathbf{H}_{n+1}(\mathbf{S}_1(\mathbf{P}_1 Y)))). \tag{A2.7}$$

The sequence of operations implied by the parenthesis leads to the following interpretation: the measurement vector Y is permuted according to $\mathbf{P}_1$ and a 0 element is affixed to the beginning of the vector. The resulting vector is Hadamard transformed by the fast algorithm. Then the first element is dropped, and the final result is obtained by permuting the vector according to $\mathbf{P}_2$ and normalizing.

For convolution, Eq. (A2.5) describes the required operation. Substituting Eq. (A2.6) into Eq. (A2.5) produces

$$Y = (\mathbf{P}_2 \mathbf{S}_2 \mathbf{H}_{n+1} \mathbf{S}_1 \mathbf{P}_1)^T F. \tag{A2.8}$$

This expression can be simplified by recognizing that $\mathbf{G}_n = \mathbf{S}_2 \mathbf{H}_{n+1} \mathbf{S}_1$ is symmetrical, giving

$$Y = [\mathbf{P}_1^T \mathbf{G}_n \mathbf{P}_2^T] F. \tag{A2.9}$$

This equation shows that only three minor changes to the original crosscorrelation program are required. Permutations are usually programmed by reducing the matrices to vectors so that the vector x can be permuted to y by the assignment

$$y(p(k)) = x(k). \tag{A2.10}$$

Because the transpose of a permutation matrix is the same as its inverse, the required operation applies the same permutation vector as follows:

$$y(k) = x(p(k)). \tag{A2.11}$$

Eq. (A2.9) also indicates that the first operation is to permute F according to the inverse of $\mathbf{P}_2$ and the last is to permute according to the inverse of $\mathbf{P}_1$—the permutations are used in reverse order. The final change is to omit the normalization. Thus, convolving noise with any filter impulse response can also be executed efficiently.

The correctness of the algorithm can be confirmed by crosscorrelating the colored noise to
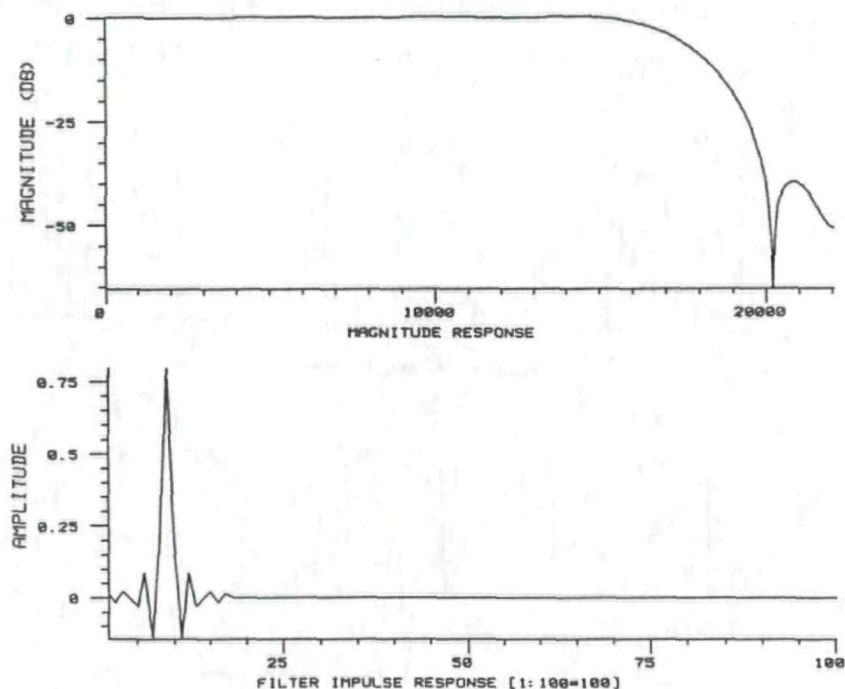
Fig. A2.1 The characteristics of the linear phase FIR filter used to demonstrate the effect of the algorithm are shown plotted in the time and frequency domains. The cutoff frequency is about 17 kHz, and the group delay is 8 samples.

see that the original filter impulse response results:

$$\Phi_{wy} = \frac{1}{n}\mathbf{W}_n Y = \frac{1}{n}\mathbf{W}_n(\mathbf{W}_n^T F) = F \qquad (A2.12)$$

because $\mathbf{W}_n\mathbf{W}_n^T = n\mathbf{I}_n$.

A2.2   DEMONSTRATION

To demonstrate the effect of this algorithm, a maximal-length sequence was filtered with a linear phase FIR filter whose impulse and frequency response are shown in Fig. A2.1. The filter characteristics were chosen to affect only the highest frequencies to preserve the overall shape. The effect of the filtering is illustrated in Fig. A2.2, where the noise sequence is plotted in the time domain before and after filtering. The filter has a group delay of 8 samples, so to maintain the correspondence the result is shown 8 samples later than the input.

The algorithm is practicable only for FIR filters, and is practical only for long impulse
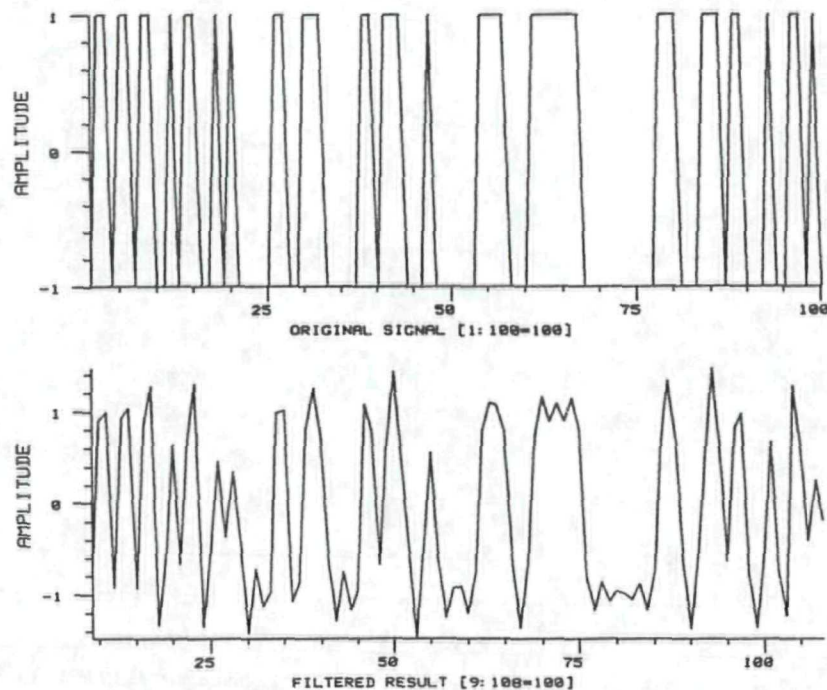
Fig. A2.2   The effect of the filtering is clear in this comparison of the noise before and after filtering. The filtered result is shown 8 samples later to allow for the group delay of the filter.

responses. The impulse response must be as long as the period of the pseudorandom noise, so a shorter impulse response must be padded with zeros. Direct convolution would be preferable when the impulse response is very short.

A2.3   CONCLUSION

Colored noise can be generated efficiently using an algorithm almost identical to the one presented previously for measuring the impulse response. As before, the algorithm is very fast because only addition and subtraction are required, and the number of operations is approximately $2.5n \log_2 n$.

APPENDIX 3

SELF-CONTAINED CROSSCORRELATION PROGRAM

A3.0 INTRODUCTION

In this appendix we present a program for performing the crosscorrelation using the algorithm described in Ch. 3. Conventionally the permutation vectors are precomputed and stored for later use during the permutation operation [NELSON-70][HARWIT-79][ALRUTZ-83]. Precomputing simplifies the programing because then the elements of the permutation do not have to be generated in the same order they will be used. More importantly, precomputing can expedite the processing, particularly when crosscorrelating many sequences with the same length. However, the results of the precomputation consume an amount of memory equal to that required for the crosscorrelation itself, so disk storage might be needed when the sequences are long. We call this approach "external" because the computation of the permutation vectors is external to the crosscorrelation routine.

The other approach generates the permutation internally, i.e., during the permutation operation. A self-contained program is more convenient to use and minimizes the amount of memory required. Of course, the execution is slightly slowed by having to compute the

permutations, but the penalty is surprisingly small because retrieving the permutations from disk partially offsets the advantage of precomputing. The internal approach excels in custom hardware because much memory can be traded for a little hardware, and the execution is not slowed at all. The tradeoffs between the internal and external approaches are comparable to the FFT where one could precompute the twiddle factors and the bit-reversal permutation, although this is seldom done.

A3.1   PERM1

As discussed in Ch. 3, permutation 1 rearranges the columns of the noise matrix $\mathbf{W}$ so that the tags are in ascending order. The tag for each column is obtained by forming a binary number from the same m consecutive terms of each column. Because of the way the matrix was generated, these m consecutive terms correspond to the successive contents of a shift register configured to produce a maximal-length sequence. The decision about which m consecutive terms to consider is equivalent to deciding the initial condition of the shift register. The choice is arbitrary as far as the first permutation is concerned because the second permutation changes appropriately so that the overall effect is always the same. We will see that choosing the first m bits simplifies the second permutation (although the last m would also be a reasonable choice).

Assuming the word-size of the computer is large enough, a single word can be used to store the present state of the shift register. The binary weight assigned to each bit is arbitrary and does not even have to be in monotonic order. But monotonic ordering is clearly the best choice because then the numerical value of the shift-register word can be used as the tag. However, there is still a question concerning the order in which the bits are assigned a binary weight— whether the first stage of the shift register is assigned to the LSB or the MSB. Usually the shift

$$
\tilde{\mathbf{W}} = \begin{array}{c} \begin{array}{ccccccc} 4 & 6 & 7 & 3 & 5 & 2 & 1 \end{array} \\ \left[\begin{array}{ccccccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \end{array} \rightarrow \tilde{\mathbf{W}}' = \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \left[\begin{array}{ccccccc} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{array}\right] \end{array}
$$

Fig. A3.1   Assigning the bit weights in the reverse order from Ch. 3 changes perm1 as shown.

```
FOR i ← 1 STEP 1 UNTIL seqLen DO
  BEGIN 'perm1'

  Permed[shiftReg] ← UnPermed[i];

  COMMENT If tap output is 1, increment shiftRegIn modulo 2;
  shiftRegIn ← 0;
  FOR j ← 1 STEP 1 UNTIL #taps DO
    IF (shiftReg LAND mask[j])
       THEN shiftRegIn ← shiftRegIn XOR xorBit;
                                        COMMENT LAND: bit-wise AND;
  COMMENT Shift right one and OR in new input at LSB;
  shiftReg ← (shiftReg ASH -1) LOR shiftRegIn;
                                        COMMENT LOR: bit-wise OR;
                                        COMMENT ASH: arithmetic shift;

  END 'perm1';
```

Fig. A3.2   The code for the inner loop of perm1. xorBit has a single one at the position of the first stage of the shift register, and shiftReg is initialized to xorBit. The array mask is used to mask out a single bit of shiftReg corresponding to the output of one of the taps.

register will be shorter than the word size, so the low-order m bits must be used. Bits that shift out of the essential m must be ignored. Left shifting, as suggested in Ch. 3, is more intuitive because the last stage of the shift register is considered the MSB. But then bits that shift out of the MSB must be gated out—an additional operation. Shifting right saves one operation because bits automatically spill out the end of the shift register. When the feedback term is 1, it is ORed into the $(m-1)$th bit position (where the LSB is defined as bit 0). By shifting right, the last stage of the shift register becomes the LSB instead, so, as shown in Fig. A3.1, this scheme produces a different permutation than the one in Ch. 3.

A3.2   PERM2

The second permutation is also straightforward to generate on the fly, although the derivation is more intricate. Because of the way $\tilde{W}'$ was created by rearranging the columns of W, m terms of each column are known—the m used to create the column tag. The row tags are determined by assembling m terms from each row, so m row tags can be deduced a priori. Using the first m rows to create the column tags predetermines the first m row tags.

The m terms selected from each row to form the row tag are the ones from columns 2^k, where $k = 0, 1, \ldots, m - 1$. Clearly, the column tag for the 2^kth column must have been 2^k.

$$\tilde{\mathbf{W}}' = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \begin{matrix} 4 \\ 2 \\ 1 \\ 6 \\ 3 \\ 7 \\ 5 \end{matrix} \quad \rightarrow \quad \tilde{\mathbf{G}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix}$$

Fig. A3.3 The second permutation changes also so that the overall effect is the same.

Therefore, the first m row tags must also be powers of 2. Whether the powers are in ascending or descending order depends on the order of significance assigned to the bits when generating the column tag. Considering the last stage of the shift register to be the LSB—as in the previous section—results in the powers falling in descending order.

The columns of $\tilde{\mathbf{W}}'$ are circular shifts of the same maximal-length sequence, so they can be computed by initializing m shift registers to the appropriate powers of 2. The row tag is assembled by assigning a binary weight of 2^k to the output of the shift register for column 2^k. Using a word to represent the state of each stage of all the shift registers (rather than one word for each shift register, as before) produces the desired weighting automatically by simply taking the numerical value of the word. Performing bit-by-bit operations extends the computation of the recursion to all the shift registers simultaneously, making the computation very efficient.

```
FOR i ← 1 STEP 1 UNTIL seqLen DO
    BEGIN "perm2"

    COMMENT Bit-wise exclusive OR of shiftReg taps forms feedback term;
    shiftRegIn ← 0;
    FOR j ← 1 STEP 1 UNTIL #taps DO
        shiftRegIn ← shiftRegIn XOR shiftReg[tapPntr[j]];

    shiftReg[tapPntr[#taps]] ← shiftRegIn;

    COMMENT Update tapPntr values by decrementing circularly;
    FOR j ← 1 STEP 1 UNTIL #taps DO tapPntr[j] ←
        IF (tapPntr[j] > 0) THEN (tapPntr[j] - 1) ELSE (shiftRegLen - 1);

    Permed[i] ← UnPerm[shiftRegIn];
    END "perm2";
```

Fig. A3.4 The code for the inner loop of perm2. shiftReg is an array whose entries represent the state of the corresponding stages of the shift registers. The shift register is realized as a circular buffer, and tapPntr is the array of pointers to the shift register taps. They are circularly decremented once each pass.

```
mMax + 1;
DO BEGIN 'fht'
    iStep + 2 * mMax;
    FOR m + 1 STEP 1 UNTIL mMax DO
        FOR i + (m - 1) STEP iStep UNTIL #samps DO
            BEGIN 'butterfly'
            j + i + mMax;
            temp + fhtArr[j];
            fhtArr[j] + fhtArr[i] - temp;
            fhtArr[i] + fhtArr[i] + temp;
            END 'butterfly';
    mMax + iStep;
    END 'fht'
UNTIL mMax ≥ #samps;
```

Fig. A3.5   The code for the FHT is adapted from a standard FFT program by changing the twiddle factors to ±1.

A3.3   FAST HADAMARD TRANSFORM

The fast Hadamard transform has the same flow diagram as the fast Fourier transform. The only difference is that the twiddle factors are all unity reflecting the fact that no multiplications are necessary. The program presented here has been adapted from the FFT routine FOUREA by Rader [RADER-79].

A3.4   DESIGNING THE SHIFT REGISTERS

To design the shift registers for generating the maximal-length sequences, one must know the appropriate recursion relation for the desired sequence length. The recursion relation is always a modulo-two sum of selected outputs of a shift register. The table in Fig. A3.6, adapted from [MACWILLIAMS-76], contains the appropriate shift register taps. Even at a sampling rate of 100 kHz, the sequence produced by a 40-stage shift register will not repeat for over 4 months, so this table should be adequate for most audio applications, although higher-order recursions are known.

To illustrate, for the case $m = 3$ the variables in the routine for the first permutation have the values #taps=2, xorBit=0000 0100, mask[1]=0000 0100, and mask[2]=0000 0001. For the second permutation the tap pointers are initialized to tapPntr[1]=1, tapPntr[2]=2, and the

m	taps	m	taps
		21	21 2
2	2 1	22	22 1
3	3 1	23	23 5
4	4 1	24	24 4 3 1
5	5 2	25	25 3
6	6 1	26	26 8 7 1
7	7 1	27	27 8 7 1
8	8 6 5 1	28	28 3
9	9 4	29	29 2
10	10 3	30	30 16 15 1
11	11 2	31	31 3
12	12 7 4 3	32	32 28 27 1
13	13 4 3 1	33	33 13
14	14 12 11 1	34	34 15 14 1
15	15 1	35	35 2
16	16 5 3 2	36	36 11
17	17 3	37	37 12 10 2
18	18 7	38	38 6 5 1
19	19 6 5 1	39	39 4
20	20 3	40	40 21 19 2

Fig. A3.6 Table showing shift-register taps to generate maximal-length sequence. The first stage is designated 1 with subsequent stages numbered consecutively.

shift register array is initialized to shiftRef[0]=1, shiftReg[1]=2, and shiftReg[2]=4.

A3.5 CONCLUSION

A self-contained program is a convenient way to implement the crosscorrelation algorithm of Ch. 3. We have found empirically that precomputing the permutations speeds the execution by a factor of only about 30% with $m = 10$. For larger m the speeds are even closer.

REFERENCES

Numbers on the first line in parenthesis are the pages on which a reference appears.

[ALLEN-79]   (6 11 12 104)
J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950 (1979).

[ALRUTZ-83]   (126)
H. Alrutz and M.R. Schroeder, "A fast Hadamard transform method for the evaluation of measurements using pseudorandom test signals," 11th Int. Conf. on Acoust., Paris, pp. 235–238 (1983).

[BAUER-61]   (59 73)
B.B. Bauer, "Stereophonic earphones and binaural loudspeakers," *J. Aud. Eng. Soc.*, vol. 9, n. 2, pp. 148–151 (1961 April).

[BARRON-71]   (78 82 85)
M. Barron, "The subjective effects of first reflections in concert halls—the need for lateral reflections," *J. Sound Vib.*, vol. 14, n. 4, pp. 475–494 (1971 April).

[BARRON-74]   (6 11 25 26 56)
M.F.E. Barron, "The effects of early reflections on subjective acoustical quality in concert halls," Ph.D. thesis, University of Southampton (1974 January).

[BARRON-81]   (83 84 114)
M. Barron and A.H. Marshall, "Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure," *J. Sound Vib.*, vol. 77, n. 2, pp. 211–232 (1981 February).

[BARRON-83]   (3)
M. Barron, "Auditorium acoustic modelling now," *Applied Acoustics*, vol. 16, pp. 279–290 (1983).

[BAXA-76]   (11)
D.E. Baxa, "A strategy for the optimum design of acoustic space," Ph.D. thesis, University of Wisconsin—Madison (1976).

[BAXA-78]   (6)
D.E. Baxa and A. Seireg, "A strategy for the optimum design of acoustic space," *Trans. ASME/J. Mech. Des.*, vol. 100, pp. 411–416 (1978 July).

[BAXA-80]   (6)
D.E. Baxa and A. Seireg, "The use of quantitative criteria for the optimum design of concert halls," *J. Acoust. Soc. Am.*, vol. 67, n. 2, pp. 2045–2054 (1980 June).

[BERANEK-54]   (49)
    L.L. Beranek, *Acoustics*, McGraw-Hill Book Company, New York, 1954.

[BERMAN-77]   (48)
    J.M. Berman and L.R. Fincham, "Application of digital techniques to the measurement
    of loudspeakers," *J. Aud. Eng. Soc.*, vol. 25, pp. 370–384 (1977 June).

[BIOT-68]   (108)
    M.A. Biot, "Generalized boundary condition for multiple scatter in acoustic reflection,"
    J. Acoust. Soc. Am., vol. 44, n. 6, pp. 1616–1622 (1968).

[BLAUERT-83]   (57)
    J. Blauert, *Spatial Hearing*, MIT Press, Cambridge, Massachusetts, 1983.

[BLIVEN-76]   (78 105)
    B. Bliven, "Annals of architecture: a better sound," *New Yorker*, vol. 52, p. 51 following,
    (1976 November 8).

[BORISH-83]   (80)
    J. Borish, "A digital delay line, part I," *The Audio Amateur*, vol. 14, n. 1, pp. 7–12 (1983
    April).

[BROADHURST-80]   (113)
    A.D. Broadhurst, "An acoustic telescope for architectural acoustic measurements,"
    Acustica, vol. 46, pp. 299–310 (1980).

[BURD-75]   (3)
    A.N. Burd, "Acoustic modelling—design tool or research project?" in R. Mackenzie
    (editor), *Auditorium Acoustics*, Applied Science Publishers, London, 1975.

[BURKE-66]   (108)
    J.E. Burke and V. Twersky, "Scattering and reflection by elliptically striated surfaces,"
    JASA, vol. 40, n. 4, pp. 883–895 (1966).

[CABOT-79]   (34)
    R.C. Cabot, "Acoustic applications of cross-correlation," presented at 64th Convention
    of AES, Novol. 2–5, 1979, preprint D-3.

[CROCHIERE-83]   (117)
    R.E. Crochiere and L.R. Rabiner, *Multirate Digital Signal Processing*, Prentice-Hall,
    Englewood Cliffs, NJ, 1983.

[DAMASKE-71]   (63 64)
    P. Damaske, "Head-related two-channel stereophony with loudspeaker reproduction." *J.
    Acoust. Soc. Am.*, vol. 50, n. 4, pp. 1109–1115 (1971).

134

[DAVIES-66] (34 36)
W.D.T. Davies, "Generation and properties of maximum-length sequences," *Control*, pp. 302–304, 364–365, 431–433 (1966 June, July, August).

[DAY-75] (3)
B. Day, "Acoustic scale modelling materials," in R. Mackenzie (editor), *Auditorium Acoustics*, Applied Science Publishers, London, 1975.

[DONATO-76] (100)
R.J. Donato, "Spherical-wave reflection from a boundary of reactive impedance using a modification of Cagniard's method," *J. Acoust. Soc. Am.*, vol. 60, n. 5, pp. 999–1002 (1976 November).

[FISCHER-83] (113)
R.E. Fischer, "Adjustable acoustics derive from two electronic systems," *Architectural Record* [entire issue devoted to Hult Center for the Performing Arts in Eugene, Oregon] (1983 May).

[FRASER-76] (53)
D. Fraser, "Array permutation by index-digit permutation," *J. of the Association for Computing Machinery*, vol. 23, pp. 298–309 (1976 April).

[FRASER-79] (53)
D. Fraser, "Optimized mass storage FFT program," in *Programs for Digital Signal Processing* (IEEE Press, NY, 1979).

[GIBBS-72] (6 11 12 23 92)
B.M. Gibbs and D.K. Jones, "A simple image method for calculating the distribution of sound pressure levels within an enclosure," *Acustica*, vol. 26, pp. 24–32 (1972).

[GRAY-77] (38)
R.M. Gray, "Toeplitz and circulant matrices: II," Technical Report No. 6504-1, Information Systems Laboratory, Stanford University (1977 April).

[GREEN-71] (81)
D.M. Green, "Temporal auditory acuity," *Psychological Review*, vol. 78, n. 6, pp. 540–551 (1971).

[HARMUTH-69] (39)
H.F. Harmuth, *Transmission of Information by Orthogonal Functions*, Springer-Verlag, NY, 1969.

[HARRIS-63] (92)
C.M. Harris, "Absorption of sound in air in the audio- frequency range," *J. Acoust. Soc. Am.*, vol. 35, n. 1, pp. 11–17 (1963 January).

[HARRIS-64] (92)
C.M. Harris and W. Tempest, "Absorption of sound in air below 1000 cps," *J. Acoust. Soc. Am.*, vol. 36, n. 12, pp. 2390–2394 (1964 December).

[HARWIT-79]   (34 35 39 126)

M. Harwit and N.J.A. Sloane, *Hadamard Transform Optics*, Academic Press, NY, 1979.

[HIRATA-79]   (6)

Y. Hirata, "Geometrical acoustics for rectangular rooms," *Acustica*, vol. 43, pp. 247–252 (1979).

[HULBERT-82]   (6)

G.M. Hulbert, D.E. Baxa, and A. Seireg, "Criterion for quantitative rating and optimum design of concert halls," *J. Acoust. Soc. Am.*, vol. 71, n. 3, pp. 619–629 (1982 March).

[JAFFE-81]   (113)

C. Jaffe, "Variable acoustics for multi-use halls using electronic reflected energy systems," presented at the 101st meeting of the Acoustical Society of America (1981 May 18).

[JANSSON-75]   (111)

E.V. Jansson, "On sound radiation of musical instruments," in *Music Room Acoustics*, Royal Swedish Academy of Music (1975 April).

[JESTEADT-77]   (82)

W. Jesteadt, C.C. Wier, and D.M. Green, "Intensity discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.*, vol. 61, n. 1, pp. 169–177 (1977 January).

[JORDAN-75]   (3)

V.L. Jordan, "Model studies with particular reference to the Sydney Opera House—the evaluation of objective tests of 'acoustics' of models and halls," in R. Mackenzie (editor), *Auditorium Acoustics*, Applied Science Publishers, London, 1975.

[KELLER-53]   (104)

J.B. Keller, "The scope of the image method," *Comm. on Pure and Appl. Math.*, vol. 6, pp. 505–512 (1953).

[KLEINER-81]   (57 61)

M. Kleiner, "Speech intelligibility in real and simulated sound fields," *Acustica*, vol. 47, n. 2, pp. 55–71 (1981).

[KLEINER-82]   (61)

private communication.

[KNUDSEN-78]   (11)

V.O. Knudsen and C.M. Harris, *Acoustical Designing in Architecture*, American Institute of Physics for the Acoustical Society of America, 1978.

[KNUTH-73]   (18)

D.E. Knuth, *The Art of Computer Programming, Vol. 1/ Fundamental Algorithms*, Addison-Wesley Publishing Company, Reading, Massachusetts, 2nd edition, 1973.

[KNUTH-81]   (34)
    D.E. Knuth, *The Art of Computer Programming, vol. 2*, Addison-Wesley Publishing
    Company, Reading, MA, 1981.

[KUTTRUFF-73]   (11)
    H. Kuttruff, *Room Acoustics*, Applied Science Publishers, London, 1973.

[MACWILLIAMS-76]   (34 35 36 130)
    F.J. MacWilliams and N.J.A. Sloane, "Pseudo-random sequences and arrays," *Proceed-
    ings of the IEEE*, vol. 64, pp. 1715–1729 (1976 December).

[MILLS-58]   (64 60)
    A.W. Mills, "On the minimum audible angle," *J. Acoust. Soc. Am.*, vol. 30, n. 4, pp.
    237–246 (1958 April).

[MORIMOTO-80]   (64 65 69 77)
    M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc.
    Jpn. (E)*, vol. 1, n. 3, pp. 167–174 (1980).

[MOORE-81]   (8)
    C.H. Moore, "Time-modulated delay system and improved reverberation simulator using
    same," *U.S. Patent 4268717* (1981 May 19).

[MOORER-79]   (8 92)
    J.A. Moorer, "About this reverberation business," *Computer Music Journal*, vol. 3, n.
    2, pp. 13–28 (1979).

[NELSON-70]   (34 36 42 126)
    E.D. Nelson and M.L. Fredman, "Hadamard Spectroscopy," *Journal of the Optical
    Society of America*, vol. 60, pp. 1664–1669 (1970 December).

[NEW YORK TIMES-70]   (1)
    (1970 July 31).

[OPPENHEIM-75]   (37)
    A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood
    Cliffs, NJ, 1975.

[PARKIN-75]   (113)
    P.H. Parkin, "Assisted resonance," in R. Mackenzie (editor), *Auditorium Acoustics*,
    Applied Science Publishers, London, 1975.

[PIERCE-81]   (96)
    A.D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*,
    McGraw-Hill Book Company, New York, 1981.

[PLOMP-80]   (6)
    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).

[RABINER-74] (39)

L.R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[RADER-70] (34)

C.M. Rader, L.R. Rabiner, and R.W. Schafer, "A fast method of generating digital random numbers," *The Bell System Technical Journal*, vol. 49, pp. 2303–2310 (1970 November).

[RADER-79] (130)

C.M. Rader, "FOUREA—a short demonstration version of the FFT," in *Programs for Digital Signal Processing*, IEEE Press, New York, 1979.

[ROCKMORE-unp] (113)

A.J. Rockmore, "Space-time signal processing," unpublished.

[SAKAMOTO-82] (63 64 65)

N. Sakamoto, T. Gotoh, T. Kogure, M. Shimbo, and A. Clegg, "Controlling sound-image localization in stereophonic reproduction," *J. Aud. Eng. Soc.*, vol. 26, n. 11, pp. 794–799 (1981 November), and part II: vol. 30, n. 10, pp. 719–722 (1982 October).

[SAKURAI-81] (100)

Y. Sakurai and K. Nagata, "Sound reflections of a rigid plane panel and of the "live end" composed by those panels," *J. Acoust. Soc. Jpn. (E)*, vol. 2, n. 1, pp. 5–14 (1981).

[SAKURAI-82] (101)

Y. Sakurai and K. Ishida, "Multiple reflections between rigid plane panels," *J. Acoust. Soc. Jpn. (E)*, vol. 3, n. 3, pp. 183–190 (1982).

[SANTON-76] (6)

F. Santon, "Numerical prediction of echograms and of the intelligibility of speech in rooms," *J. Acoust. Soc. Am.*, vol. 59, n. 6, pp. 1399–1405 (1976 June).

[SARWATE-80] (35)

D.V. Sarwate and M.B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences," it Proceedings of the IEEE, vol. 68, pp. 593–619 (1980 May).

[SCHROEDER-61] (8)

M.R. Schroeder and B.F. Logan, "'Colorless' artificial reverberation," *J. Aud. Eng. Soc.*, vol. 9, n. 3, pp. 192–197 (1961 July).

[SCHROEDER-62] (8)

M.R. Schroeder, "Natural sounding artificial reverberation," *J. Aud. Eng. Soc.*, vol. 10, n. 3, pp. 219–223 (1962 July).

[SCHROEDER-63]   (64)

M.R. Schroeder and B.S. Atal, "Computer simulation of sound transmission in room," IEEE Intern. Conv. Rec., vol. 11, pp. 150–155 (1963).

[SCHROEDER-70]   (6 7 28 30 59 63 64 73)

M.R. Schroeder, "Digital simulation of sound transmission in reverberant spaces," J. Acoust. Soc. Am., vol. 47, n. 2, pp. 424–431 (1970).

[SCHROEDER-73]   (63)

M.R. Schroeder, "Computer models for concert hall acoustics," Am. J. Physics, vol. 41, pp. 461–471 (1973 April).

[SCHROEDER-79a]   (78 105)

M.R. Schroeder, "Binaural dissimilarity and optimum ceilings for concert halls: more lateral sound diffusion," J. Acoust. Soc. Am., vol. 65, n. 4, pp. 958–963 (1979 April).

[SCHROEDER-79b]   (34 48)

M.R. Schroeder, "Integrated-impulse method measuring sound decay without using impulses," J. Acoust. Soc. Am., vol. 66, pp. 497–500 (1979 August).

[THOMASSON-76]   (100)

S. Thomasson, "Reflection of waves from a point source by an impedance boundary," J. Acoust. Soc. Am., vol. 59, n. 4, pp. 780–785 (1976 November).

[TOLSTOY-79]   (108)

I. Tolstoy, "The scattering of spherical pulses by slightly rough surfaces," J. Acoust. Soc. Am., vol. 66, n. 4, pp. 1135–1144 (1979 October).

[WALSH-81]   (7 28 30)

J.P. Walsh and N. Dadoun, "The design and development of Godot: a system for computer-aided room acoustics modeling and simulation," 101st meeting Acoust. Soc. Am., Ottawa, Canada, paper P7 (1981 May).

[WALSH-82]   (7)

J.P. Walsh and M.T. Rivard, "Signal processing aspects of Godot: a system for computer-aided room acoustics modeling and simulation," 72nd convention Aud. Eng. Soc., Anaheim, CA, paper A-10 (1982 October 23–27).

[WATSON-83a]   (108)

J.G. Watson and J.B. Keller, "Reflection, scattering, and absorption of acoustic waves by rough surfaces," J. Acoust. Soc. Am., vol. 74, n. 6, pp. 1887–1894 (1983 December).

[WATSON-83b]   (108)

J.G. Watson, D. Stickler, and J.B. Keller, "Reflection, transmission and scattering by rough interfaces," unpublished (1983).

[WAYMAN-77]   (6)

J.L. Wayman and J.P. Vanyo, "Three-dimensional computer simulation of reverberation in an enclosure," J. Acoust. Soc. Am., vol. 62, n. 1, pp. 213–215 (1977 July).

[WAYMAN-80]   (28 30)
    J.L. Wayman, "Computer simulation of sound fields using ray methods," Ph.D. thesis,
    University of California—Santa Barbara. (1980 July).

[WENZEL-74]   (100)
    A.R. Wenzel, "Propagation of waves along an impedance boundary," *J. Acoust. Soc.
    Am.*, vol. 55, n. 5, pp. 956–963 (1974 May).