CLARITY, COCKTAILS, AND CONCERTS: LISTENING IN CONCERT HALLS

David Griesinger
221 Mt. Auburn Street, #107
Cambridge, Massachusetts 02138

How can we decode the complexities of music with only two ears?

The key to understanding musical acoustics lies in the extraordinary ability of the human ear and brain to extract a wealth of precise information from a complex and often chaotic sound field. A human ear has only about 3,500 sound-sensing hair cells, each capable of firing no faster than 1000 times per second. They are attached to a frequency sensitive mechanical filter with a selectivity of about one part in five. Yet with this meager data we can tune instruments to one part in a thousand, choose to listen to any one of several simultaneous conversations in a noisy room (the cocktail party effect), or know which instrument played each note in a string quartet.

"The ability to separate individual sources from a complex sound field has limits."

The ability to separate individual sources from a complex sound field has limits. When noise and reverberation are too strong the ability to hear more than one conversation or more than one musical line vanishes. We can no longer perceive all the complexity of music and our attention is more likely to wander. This article uses clues from physics and our perception of music to understand these limits, and how they influence concert hall design.

Sorting sound waves into sound events such as notes and syllables and then assembling these events into coherent streams with similar pitch, timbre, location, and distance is the job of specialized organs in the brain stem—the oldest part of our neurology. Much has been learned about the function of these organs with animals, and much is still mysterious. But our ability to hear music gives powerful clues to how the mechanisms work. From these clues we can begin to understand the physics of the process; how information about timbre and localization is encoded in sound waves, how the ear and brain extract this information, and how, and to what extent, reflections and noise interfere.

The brain stem works at a subconscious level. The process of sorting sound into many simultaneous foreground and background sound streams that can be assembled into a meaningful image is automatic; we cannot influence it by thought. The streams are passed upward to consciousness fully formed. There they are processed in ways well beyond the scope of this article. But with music and physics as our guide we can start to make sense of the processes going on in the brain's subconscious realm.

Vision, hearing, and sound streams

Human perception is multi-modal. The brain makes sense of reality by combining information from many senses. In a music performance we hear the sound of an instrument at the direction and distance we see it regardless of our ears. Improving the lighting improves the clarity of the sound, and the color of a hall changes the perception of sound dramatically. The ticket price nearly always reflects the clarity of vision, not sound, and new halls with spectacular architecture sound wonderful for a while. But while we can hear many things with our eyes, what we do or don't hear with our ears affects us profoundly.

In good acoustic spaces if we close our eyes for five minutes or more and shake our heads a few times we can still perceive the pitch, timbre, direction and distance of multiple sound sources at the same time. Somehow our ears and brain stem have managed to separate a jumble of overlapping sounds into separate streams of information, one for each source. This is the well-known cocktail party effect, vital to our survival as a social species. When it is possible to separate sounds in this way the brain stem sends multiple streams upward to consciousness. When it is not possible we hear the whole ensemble as a single mixed stream of sound, and are unable to localize or identify the timbre of individual voices. For speech the result is babble. Music is more forgiving. We hear pitches, harmony, loudness and dynamics—but information about who played what, from where, and with what timbre, is lost.

The difference is not subtle. When we can separate voices the higher brain is able to process the sound with far greater attention and interest. Imagine being at a party so crowded that you cannot hear what anyone is saying, and contrast that with the situation where we are able to listen to any one of several conversations at will. The first situation will cause the brain to tune out, the second demands attention.

Direct sound—The key to localization and timbre

Sound in open air decreases in level by six decibels for each doubling of distance. People sitting close to the source will hear a much louder sound than people sitting further away. One fundamental purpose of a hall is to catch sound that does not go directly to the listeners and re-direct it toward them. This makes the sound louder and more uniform. In halls very few of the reflections that re-direct the sound are first-order (i.e., have bounced off only one surface before being heard), and the first order reflections are always weaker individually than the sound that travels directly to the listener. But there are a great many reflections of higher order, and they combine chaotically to create what we call reverberation. In a typical seat in a concert hall, the combination of reflections and reverberation contains at least ten times more energy than the direct sound, and most of that energy comes from high order reflections. This creates the

DOWNLOAD THE AUDIO CLIP EXAMPLES

Shortly after the print copy of this issue is mailed, it will also be published in the Acoustical Society of America's Digital Library. The Table of Contents may be reached directly by going to your internet browser and typing the following Uniform Resource Locator (URL) in the address block: http://scitation.aip.org/dbt/dbt.jsp?KEY=ATCODK&Volume=7&Issue=1. At some point in the download process you will probably be asked for your ID and Password if you are using a computer that is not networked to an institution with a subscription to the ASA Digital Library. These are the same username and password that you use as an ASA member to access the *Journal of the Acoustical Society of America*. Open the abstract for this article by clicking on Abstract. At the bottom of the abstract will be a link. Click on the link and a zipped folder (DavidGriesinger.zip) will be downloaded to your computer (wherever you usually receive downlinks). Unzip the entire folder (as a whole), save the unzipped folder to your desktop. There are 2 examples. Do not remove any file—simply open the file called 00_OpenDavidGriesinger. It is an interactive file. By clicking on any of the hyperlink clips in this file, you should hear the audio clip associated with that link. The best way to enjoy article is to open the magazine. Open the file on your computer. With mouse in hand, read the article from the magazine and when a example is mentioned, click on the link associated with the clip. Good PC speakers or earphones will enhance your perception of the subtleties present in the examples. The audio clips were recorded in .mp3 format. If you have difficulty in playing them, you might download the PC or MAC version of *VLC Media Player* from www.videolan.org. This is a non-profit organization that has created a very powerful, cross-platform, free and open source player that works with almost all video and audio formats. Questions? Email the Scitation Help Desk at help@scitation.org or call 1-800-874-6383.

loudness we seek, along with the envelopment we find so desirable.

Reverberation can be beautiful, and at least in a large hall. "louder is better." But reverberation is by nature chaotic and blending. Information about which instrument played what, and from where, is lost. For some people this is OK. They have no desire to hear every instrument with the clarity of the score, and a rich blend of all the instruments is just fine. Other people want the kind of clarity that enables you to distinctly hear the way the composer has written each line, how it is played, and from where. Even without consciously listening for these details the brain reacts more strongly to the music when this clarity is present. Clear sound, sound perceived as close to the listener, demands attention. Attention creates drama—and drama in music is addictive. Surprisingly, this kind of clarity can co-exist with reverberation and envelopment, even when the reverberation is much stronger than the direct sound.

We need to appreciate that the direct sound—the brief segment of sound that arrives at a listener before being augmented by reflections—conveys most of the information about localization and timbre. We also need to understand that in most seats in most halls the direct sound is weaker than the sum of the early reflections and reverberation that quickly overtake it. But we propose that if the auditory nerve firings from the direct sound are more numerous than the nerve firings from the reflections in the first hundred milliseconds after the beginning of a note the brain stem can create separate neural streams for each musical line, and identify which instrument played them. How can we test this proposal? How can we find the distance in a hall where the ability to separate individual voices vanishes?

Binaural recordings of the eardrum pressure

Studying sound perception in halls is difficult because the brain suppresses the conscious perception of noise, reflections, and reverberation. Thus the sound quality in a hall is difficult to judge, and almost impossible to remember. The dominance of our vision further complicates the situation. If we see musicians playing we will perceive their sounds coming from the direction we see them—regardless of whether sonic localization is possible or not. For many people it takes practice to perceive a scene from sonic information alone. But differences can be startling when the sonic images from two different halls or two different seats in the same hall are rapidly compared in the absence of a visual image.

It is possible to make binaural recordings of the sound at a listener's eardrums, and to reproduce it through headphones also equalized at the eardrums. The result is nearly perfect reproduction of an auditory scene. Surprisingly recordings made at my own eardrums are convincingly realistic for at least 50% of listeners, even without individual headphone equalization. They are particularly successful for people such as recording engineers who are accustomed to work without a visual image.

With the help of these recordings, we find that in all halls the location and timbre of individual instruments can be clearly identified when the listener (or the binaural recording position) is close to the musicians. As the listener moves back into the hall the location and timbre of individual instruments, and the ability to clearly hear their musical lines in the presence of all the other instruments, continues to be good up to a certain point. At this critical point the sound changes. Instead of perceiving a coherent image—where each instrument can be localized and identified—all the instruments blend together into a fuzzy ball of sound. Occasionally a solo instrument will be localizable, but when instruments play together they all fall into the same sonic blob. Timbre also changes dramatically. When instruments are localizable each timbre is distinctly perceived. When they are not localizable the whole ensemble takes on a darker color—one that sound engineers call "muddy". This change in timbre is distinctly perceived even when these binaural recordings are played through loudspeakers.[1]

These binaural recordings illuminate an aspect of hall acoustics that acoustic research has largely ignored. In standard texts, clarity has been loosely defined by the intelligibility of single voices, or the ability to hear the pitches of instruments. For many seats in modern concert halls and opera houses the clarity that enables a trained listener to identify and localize the instruments in an orchestra or a string quartet, the clarity that pulls the full attention of a listener into the composition, the clarity that nearly every commercial sound recording delivers, is lost.

Localization and timbre

Timbre of an instrument—and the difference between spoken vowels—is determined by the strength of harmonics in the vocal formant frequency range, roughly 700Hz to 4000Hz. The basilar membrane filters in the inner ear separate these frequencies into about 15 overlapping bands. The differences in the strength of the signal in these bands allow us to identify the word or the instrument. Likewise, differences in the strength and timing of the signals between the two ears allow us to determine the sound direction. But if several instruments are playing at once typically two or more harmonics from each source occupy the same basilar membrane filter. The basilar membrane is not selective enough to separate them. If we look at the average signal in each filter band we will get a mixture of timbres—and have little clue to the source directions.

Separation of sound sources by pitch

A critical issue for music and speech perception is that instruments playing together, or several people talking at once, all produce harmonics in the same vocal formant range. If we are to detect the location and timbre of each instrument or the vowels of simultaneous speech we must first separate the harmonics from each source into independent neural streams. It is clear that the brain stem can do this, and the ability is vital to human hearing. The ability to separate harmonics enables us to listen to several conversations at once and switch our attention between them at will. The cocktail party effect is known to depend critically on pitch. A person speaking in a monotone can be separated from another if the difference in pitch is only half a semi-tone, a frequency difference of only three percent. If the pitches are identical—or if the speakers whisper—the two voices cannot be separated. We believe that the necessity of performing the cocktail party effect has driven the evolution of our extraordinary sensitivity to pitch—and of our appreciation of musical scales and harmony.

The properties of music can be used to understand the physics of this process. A trained musician can tune an instrument to an accuracy of one part in a thousand. The average music lover can perceive pitches to at least 1%. The basilar membrane is incapable of such precision. Furthermore, our ability to perceive pitch is circular in octaves. If we double the frequency of a complex tone, the pitch—in a musical sense—remains the same. It is sometimes difficult to decide in which octave a complex tone originates, particularly in the presence of other pitches.

The author has developed a physical model that explains these abilities. Physics tells us that harmonics carry in their phase the memory of the pulse that created them. If several adjacent harmonics of the same tone are present at the output of a filter, once in each fundamental period the harmonics align in phase, adding together to make a strong peak in the output of the filter. As the harmonics drift apart the peak goes down. The result is a strong amplitude modulation of the filter output. When several harmonic tones are present at the same time each creates modulations specific to their fundamental frequency and these modulations sum linearly. In this model the basilar membrane is not only sensitive to the average amplitude in a band, but it also detects amplitude modulations in that band—much like an AM radio.

In our model the detected modulations from each band pass to a group of neural structures that resemble comb filters—a pitch sensitive filter that is both highly efficient of neurons and circular in pitch. A comb filter can be understood as a delay line with a large number of taps, each separated by a constant delay. The output consists of the sum of

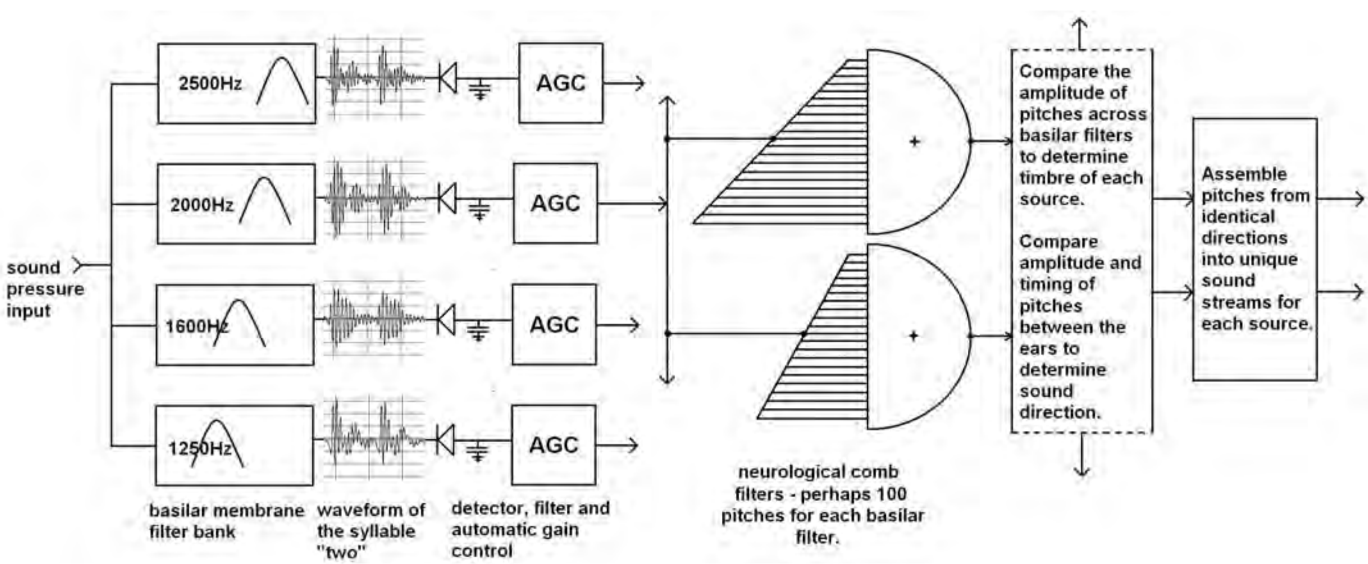


Fig. 1. Flow of information through the model.

all the taps. When the delay between each tap corresponds to the period of a particular frequency the nerve pulses at the output will sum to a high value, implying that the modulation in firing rate from that filter will be a maximum. When the tap period does not correspond to a multiple of the input frequency the output is minimal.

There are enough comb filters in each group to sort incoming modulations by their pitch into separate neural paths, one path for each pitch. To achieve the pitch accuracy of a musician, the group requires only about a hundred different comb filters, each with a total delay of about 100 milliseconds. Brief signals produce useful pitches in a fraction of that time. Figure 1 shows the flow of information through the system, and a possible neural implementation of a comb filter based on the speed of pulses traveling through fine diameter nerve fibers. The diagram in Fig. 1 shows the same number of taps for each pitch, and a variable total delay length. Our computer model uses a constant total delay for all pitches, and varies the number of taps. Which system (if any) is actually used is not predicted by our data—but the average length of the total delay must be about 100ms to match our abilities to perceive music.

In Fig. 1, sounds entering the ear are separated into frequency bands by a bank of overlapping mechanical filters with relatively low selectivity. At the vocal formant frequencies each filter typically contains three or more harmonics of speech or musical fundamentals. These harmonics interfere with each other to create a strongly amplitude modulated signal, as can be seen in the figure. The modulations in the signal are detected linearly by the hair cells, but like an AM radio with automatic gain control the nerve firing rate for time variations longer than about 20 milliseconds is approximately logarithmically proportional to the sound pressure. The brain stem separates these modulations by pitch using a number of comb filters each ~100ms long. Two of these filters (detecting two different pitches) are shown in the figure, but about one hundred are needed for each basilar membrane band. Once separated by pitch the brain stem compares the amplitude of the modulations for each pitch across the basilar filter bands to determine the timbre of the source, and compares the amplitude and timing of the modulations at each pitch between the two ears to determine sound direction. Using these cues the brain stem assembles events into separate foreground sound streams, one for each source. Sound left over after the foreground is extracted is assigned to a background sound stream. Reflections and reverberation randomize the phases of the harmonics. When the reflections are too strong the modulations in each frequency band become noise-like, and although pitch is still detectable, timbre and direction are not.

Stream formation

The comb filters separate sound events by pitch relatively easily, and can do it in the presence of high levels of reverberation. But to create separate sound streams for each source the brain stem must determine to which sound source the various pitch events belong. The task is easy if the timbre and azimuth of each pitch event can be identified, and this is possible when the acoustics are sufficiently clear. By comparing the strength of the modulations at a specific pitch across the formant bands the timbre of a particular event can be determined, and by comparing the strength and timing of each pitch event between the two ears the localization can also be determined. Using these cues the brain stem can assemble events into meaningful foreground streams, and present the streams to higher levels of the brain.

In this case the brain is capable of a further separation. Sound elements identified by their pitch, localization, and timbre can be separated from the reverberation they induce. We get a distinct perception of two different types of sonic streams—the foreground streams of notes and syllables, and a single combined background stream that includes all the reverberation. The background stream has interesting properties. When the foreground is strong the notes and syllables mask the reverberation, but we perceive the reverberation as continuing unbroken through the foreground sound events. When the reverberation is stronger than the foreground elements, the foreground elements are perceived with the timbre and azimuth that is detected at their onsets—even if the reverberation soon overwhelms them. In both cases if the background stream is at least partially coming from all directions it is perceived as surrounding the listener.

When the azimuth and timbre of the direct sound is masked by reflections and reverberation, the brain is forced to consider both the note and its reverberation as one sound event. The combination becomes one sonic object. The reverberation is bound to the note, and is perceived as primarily in front of the listener, regardless of the actual spatial distribution of the reverberation. When the foreground—the direct sound—is clearly perceived, the reverberation can be separated from the note. Then for most people the reverberation is perceived as louder and more enveloping.

But there is another aspect of stream formation. When the brain is able to accurately separate notes or syllables by pitch we perceive the instruments or speakers as being close to us. These sounds demand more attention than sounds perceived as muddy and far away. This kind of clarity is an essential part of drama and cinema. Drama and cinema directors demand that theaters be acoustically dry, with directional loudspeakers for dialog. They want the maximum dramatic effect to be conveyed to the audience. The author firmly believes the same kind of clarity is needed in musical performances. Opera especially needs clarity, whether you understand the language or not. Clear sound draws a listener into the emotional experience of the scene. Well blended sound encourages a passive kind of listening. The goal in theaters is to make the direct sound stronger than the total reverberation; to make the direct to reverberant ratio (D/R) greater than unity. But some concert halls and opera houses demonstrate that dramatic clarity can be achieved at lower values of D/R.

Implications for musical acoustics

The physical model and the observations above need not be precisely accurate to be useful for room acoustics. The physics on which they are based predicts reasons why some halls deliver startling clarity over a wide range of seats—and why many of their copies do not. First, the model explains the observation that the ear and brain can detect localization and timbre in a reverberant field more easily at the vocal formant frequencies than at the fundamental frequencies of most instruments. At low frequencies there are too few cycles in the brief time before reverberation overwhelms the direct sound. In addition the ability to separate sources into independent streams depends in part on the presence of multiple harmonics from the same source in each critical band of the basilar membrane—and this happens largely at higher frequencies. So if we can maximize the strength of the direct sound relative to the reflections and reverberation at high frequencies—while leaving the reflections strong at lower frequencies—we can achieve both good clarity and rich reverberation at the same time.

Second, the model predicts that clarity—source separation—depends on the time delay between the onset of the direct sound and the cumulative sum of reflections in a 100ms window. The larger the time delay the greater will be the sum of nerve firings from the direct sound compared to the number from the reverberation. These predictions lead to a method of understanding the clarity of halls, and why their properties do not scale with size.

Listening in concert halls

It is widely believed that a shoebox shape is ideal for classical music performance, regardless of the size of the hall and the type of music performed. Since the eye is tolerant of scale—a shoebox holds 2000 people as easily as it holds shoes—we assume the same holds for sound. But there are far more mediocre shoebox halls than great ones, and the smaller the hall, the poorer they are likely to be. In Leo Beranek's surveys of musicians and conductors only three halls are rated "excellent." Many people consider the Boston Symphony Hall (BSH), with 2625 seats, to be the best. But many of its close copies fall short. The odds of building an excellent new hall with a shoebox shape do not appear to be good—especially if the copy is smaller.

Other options exist. Currently "vineyard" halls are popular. These halls are typically oval in plan, with no overhanging balconies. The audience surrounds the orchestra in terraces (vineyards), the walls of which are intended to supply early reflections. The average listener is closer to the musicians than a typical shoebox of the same capacity, but many sit in poor seats behind the orchestra. Late reverberation in vineyard halls tends to be weak because the direct sound is either absorbed by the audience, or is directed down into the audience by panels on the ceiling (and is thus absorbed). There is little sound left over to create late reverberation. These halls lack the warmth and envelopment of BSH. A better option is exemplified by the Teatro Colón in Buenos Aires, which resembles a large semi-circular opera house. It is renowned as a concert hall where music is heard with extraordinary clarity and reverberation in a great majority of seats. But Beranek lists it in his books as an opera theater, and does not rank it as a concert hall. Jordan Hall at New England Conservatory, with 1013 seats, is a Mecca for chamber musicians and audiences from all over the world, and is also excellent for small orchestras.

Neither of these halls is a shoebox. Both are semi-circles with high balconies. They bring the average listener closer to the musicians than in a shoebox, and their high ceilings provide the cubic volume and the reflecting surfaces needed for fine late reverberation. But citing the success of BSH the public, architects, and large donors usually demand a shoebox shape. BSH beats the shoebox odds by a rare combination of many elements, starting with size, shape, and surface. Changing any of these elements makes success unlikely.

Binaural recordings of live concerts in BSH reveal that the timbre and localization of each instrument is excellent everywhere on the floor—up to about row Z, just beyond the cross aisle, ~80 feet from the conductor. At the same time the reverberation is nearly always audible as the music is playing, lending a wonderful ambience to the sound. Surprisingly both clarity and reverberance are excellent—perhaps even better than on the floor—in the front row of the first balcony, 110 feet from the conductor. BSH succeeds in delivering both clarity and reverberation over a large majority of seats—a feat extremely rare in concert halls. How does it do it?

In most seats in BSH you can (with practice) hear all the notes in the music separately, and tell which instruments played them. You can also hear the reverberation of the hall as separate from the foreground notes. To hear all the notes separately your brain must be able to process the sound that comes directly from the instruments. But the direct sound is easily muddled by excess early reflections. Our neural model shows that the brain needs about a tenth of a second of direct sound to detect the pitch, timbre, direction, and distance of each player or section as separate from the others. If the number of nerve firings from the direct sound at the start of each note is greater than the number of nerve firings from the reflections for this critical tenth of a second, the brain can detect all the information we need. If the number from the reflections is greater in this tenth of a second the instruments blend together. The reverberation becomes the notes—audible as chords, but not as separate sources. We still hear harmony, but not the notes that create the harmony.

The kind of enveloping reverberation that gives a hall its richness is only audible if the direct sound is separately detected, and then only after the critical tenth of a second has elapsed. If the majority of the reverberant energy has decayed before this time the hall will be perceived as dry. In brief—we only hear all the notes if the number of nerve firings from the direct sound is greater than the number of nerve firings from all the reflections in the first tenth of a second, and we only hear the reverberation as separate if it is still strong enough to be heard after this period has elapsed.

The success of BSH is due to two factors: the audibility of the direct sound at frequencies above 1000Hz, and the relatively high strength of the late reverberation—sound that arrives at the listener more than 100ms after the direct sound. A major tenent of current acoustic design is that strong reflections from the side of the hall are essential for good sound. The experiments on which this tenent is based assumed that the direct sound was always audible, and only tested cases where the direct sound was nearly as strong as or stronger than the sum of all the reflections. Unfortunately in almost every seat in a shoebox hall the opposite is true, and for seats more than half-way back the early reflections from any direction disturb our ability to hear the music. Vienna's renowned Grosser Musikvereinssaal nearly falls into this category. If you luckily get a ticket in the front half the sound is fantastic, but more than half-way back the sound is just loud. You are better off in the standing room. Shielded by the balcony from most of the reflections the sound is clear, well balanced, and beautiful. (But get there early. You need to be in the front of the crowd.) Unfortunately most modern hall designs direct first order reflections down into the audience, particularly from the side walls and balcony fronts. The result is disastrous. In far too many seats the direct sound is not detectable. These first-order reflections are then absorbed—and their energy cannot contribute to desirable late reverberation.

If we measure the impulse response of BSH when the stage and hall are fully occupied we find that although the reverberation time varies very little with frequency, the strength of the early reflections above 1000Hz in the rear half of the hall is weaker than it is at lower frequencies. The front of the first balcony is especially blessed by this lower level of early reflections. The reason the first balcony succeeds is clear from the geometry of the hall. This position does not receive strong early reflections from any direction. The underside of the first balcony along the side walls directs all the specular lateral reflections that would otherwise travel to the first balcony down into the seats on the floor. This explains why listeners beyond row Z have difficulty localizing. The balcony fronts are transparent to sound, and have absorbing legs of audience behind them. The strong early reflection that normally comes from the ceiling is deflected back to the orchestra by the many coffers that decorate the ceiling. The coffers in BSH have just the dimensions needed to act as retro-reflectors for frequencies above 1000Hz, while allowing lower frequencies to reflect specularly. The absence of strong early reflections in the front of the first balcony gives the brain stem time to separate the direct sound. The result is high clarity and rich reverberation.

Most shoebox halls fail to provide the clarity and reverberation of BSH because the early reflections in the rear of the hall are too strong, and come too soon. Seats in the front of such halls are not problematic, as the direct sound is strong, the early reflections are relatively weak, and they have a longer delay relative to the direct sound. Knowledgeable people and critics sit there. As you move back in the hall the direct sound is weaker, the first-order reflections are stronger, and they come sooner. At some critical distance localization becomes impossible, and the instruments blend into a circular blob of sound. Our experiments and binaural recordings show that the boundary between the two types of sound is often only one or two meters wide. Seats in front of this critical distance give wonderful, and nearly identical, sound. Beyond this critical distance for localization the sound is muddy and blended. In countless halls most of the seats have this kind of sound.

There is a simple graphic that lets you see how the ear is hearing the beginning of a sound event. Let's assume we have a sound source that suddenly turns on and then holds a constant level. Initially only the direct sound stimulates the basilar membrane. Soon the first reflection joins it, and then the next, etc. The nerve firing rate from the combination of sounds is approximately proportional to the logarithm of the total sound pressure. But instead of plotting the total rate of nerve firings we plot the rate of nerve firings from the direct sound and the reflections separately. In the following graphs the vertical axis is labeled "rate of nerve firings", normalized such that the rate is 20 units for the sum of both rates once the reverberation is fully built-up. The scale is chosen such that the value of the rate can be interpreted as proportional to the decibels of sound pressure. Thus in Fig. 2, the rate for the direct sound is about 13, implying that the total sound pressure will eventually be 7dB stronger than the direct sound. Figure 2 shows the relative rate of nerve firings from the direct sound and the build-up of reverberation in the frequency range of 1000Hz to 4000Hz in unoccupied Boston Symphony Hall (BSH) row R, seat 11, with a source at the podium. The dashed line shows the rate of nerve firings for a sound of constant level that begins at time zero. The solid line shows the firing rate due to the reverberation as it builds up with time. The dotted line marks the combined final firing rate for a continuous excitation, and the 100 ms length of the time window the brain stem uses to detect the direct sound. In this seat the direct sound is strong enough that the ratio of the area in the window under the direct sound (the total number of nerve firings from the direct sound in this window) to the area in the window under the build-up of the reflections is 5.5dB. This implies excellent localization and clarity. Shown in Fig. 3 are the nerve firing rates for the direct sound and the build-up of reflections in unoccupied BSH, row DD, seat 11. Notice the direct sound is weaker than in row R, and there is a strong high-level reflection at about 17ms that causes the reflected energy to build up quickly. The ratio of the areas (the total number of nerve firings) for the direct sound in the first 100ms to the area under the line showing the build-up of the reflections is 1.5dB. Localization is poor in this seat.

Figure 4 shows the graph for the front of the first balcony.

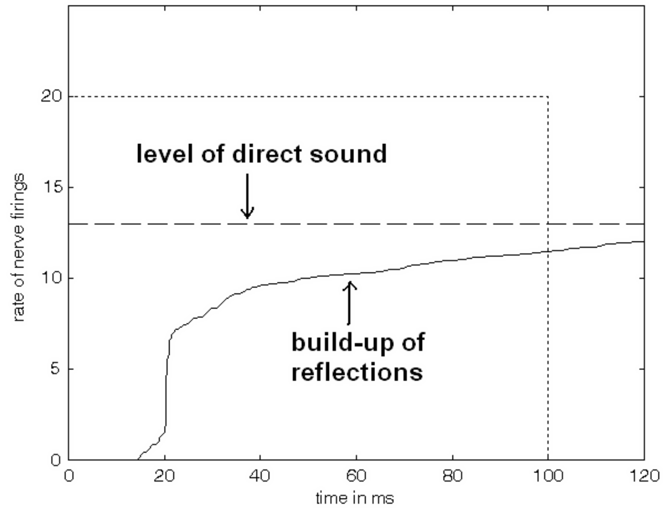


Fig. 2. The relative rate of nerve firings from the direct sound and the build-up of reverberation. in the frequency range of 1000Hz to 4000Hz in unoccupied Boston Symphony Hall, row R, seat 11, with a source at the podium.

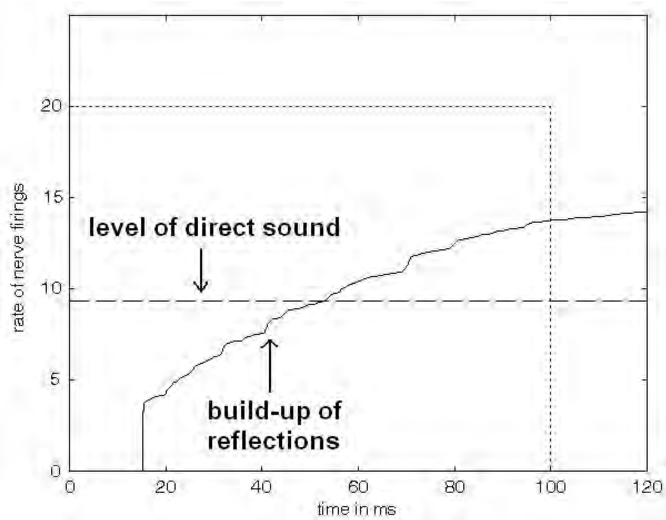


Fig. 3. Nerve firing rates for the direct sound and the build-up of reflections in unoccupied Boston Symphony Hall, row DD, seat 11.

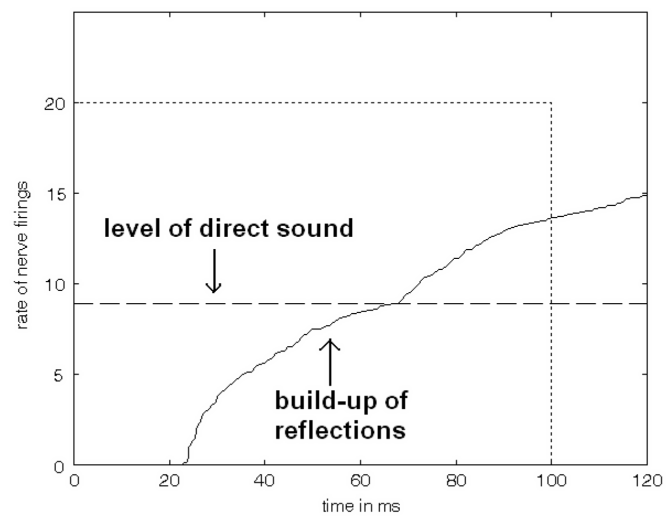


Fig. 4. Rates of nerve firings for the direct sound and build-up of reflections in Boston Symphony Hall, front of first balcony, row A, seat 23.

The direct sound is weaker here—but there are no strong early reflections. The ratio of areas is +2.2dB, and localization is better than in row DD on the floor. (See Fig. 3) The localizability predicted by the ratio of areas is poorer than my subjective impression in the fully occupied hall—but the difference in the way the reflections build up is easy to see.

Mitigation

Halls need not sound either muddy or too dry. Some of the old shoebox halls, and almost all of the new ones, lack the coffers and niches that make BSH work. But it is possible to add elements that perform the same job. Plastic saucers or cloud elements of variable size over the orchestra can be arranged to reflect frequencies above 1000 Hz down into the orchestra and the front rows of the audience, while letting lower frequencies excite the upper volume of the hall. The high frequencies will be absorbed, increasing the high frequency D/R in the rear of the hall without changing the reverberation time. Clarity in the rear will improve. The direct sound is strong in the front, and the prompt early reflections will be appreciated. Beams and columns added to the side walls perform the same function—namely they reflect the high frequency portion of the lateral reflections back to the front, giving people in the rear more time to detect the direct sound.

But one needs to be careful. It is possible to reduce the early reflections too much, or make a hall too wide. There needs to be enough energy between the direct sound and the bulk of the reverberation to prevent the brain stem from detecting the reverberation as a separate foreground event—an echo. The marvelous Concertgebouw in Amsterdam, with substantially greater width than BSH, is at the limit. Orchestral music is gorgeous. The sound is both clearer than BSH and more reverberant. But in some seats during a piano performance the reverberation is heard disconnected from the notes.

In smaller halls all the reflections come sooner, and the reverberation builds up more quickly. In small halls the sound is loud and muddy almost everywhere, especially with student orchestras playing modern instruments. Such halls are also perceived as too dry, as the volume is not large enough to allow a strong late reverberation. The owners of such halls are very reluctant to add absorption, as this will make lower the reverberation time. But adding absorption to the stage can be surprisingly effective. Vaudeville stages

Search for Low Frequency Underwater Seismic Sources

Traditionally, the marine seismic prospecting industry uses airgun impulsive sources for generating acoustic signals in water that will be reflected back from rock formations thousands of meters below. The Global Petroleum Research Institute (GPRI), affiliated to Texas A&M University and sponsored by a group of major oil companies, is searching for alternative technologies capable of transmitting a lower power, longer duration signal of equivalent energy. Such technology may exist in applications such as Low Frequency Sonar and Oceanography. The purpose is to identify a few promising technologies, to fund the building or adaptation of prototypes to meet seismic specifications, and to verify the reliability and far field acoustic output. The technology will then be made commercially available to the seismic prospecting industry.

Key Seismic Specifications

Array output for 5 s signal

- 5-10 Hz 190 dB re 1 uPa/Hz @ 1 m
- 10-100 Hz 200 dB re 1 uPa/Hz @ 1 m

Harmonic content above 150 Hz

- 40 dB down when driven by tone in 5-100 Hz range

Reliability

- 72 sweep hours between maintenance
- 720 sweep hours between overhaul

If you have a technology for consideration, please contact Mr. David Burnett, GPRI – Texas A&M University JIP Representative. Please visit www.gpri.org for contact details.

almost invariably had curtains in the wings, at the back of the stage, and above the proscenium. Sound from the performers that did not travel directly to the audience was absorbed, increasing the relative strength of the direct sound. Excess loudness was also controlled—and the drama of the performance was maximized. We need not use curtains or completely cover the stage walls to get the benefit.

The bottom line: we need to adjust the shape of a hall to match the size of the hall. Large halls can use a shoebox shape successfully if they absorb or back scatter the first order reflections that would otherwise travel from the orchestra to the rear of the hall. This suggestion may be anathema to those acousticians who believe these reflections are essential to give support and loudness to the listeners in the rear. But these reflections contribute very little to loudness—remember that individually they are weaker than the direct sound, and the total reverberant energy is often more than ten times the strength of the direct sound. Loudness comes from the reverberation. Sitting in the first row of the first balcony in BSH proves the point.

As halls become smaller the design goal should be to choose a shape that brings the audience closer to the musicians, and to obtain the needed reverberation by increasing the room volume overhead. Such halls need not have the 1.9 second reverberation time of BSH. BSH is large enough provide late reverberation without excessive energy in the first 100 milliseconds of decay. Efforts to reduce the total absorption of a small hall to the point where it can achieve the same reverberation time as BSH will result in massive amounts of early reverberant energy. This will prevent a clear sound, and prevent the formation of a background sound stream. The hall will sound less reverberant than if the sound was clear and the reverberation time was only 1.3 or 1.4 seconds. Jordan Hall at New England Conservatory demonstrates this effect beautifully.

There are examples of very small halls (~300 seats) that manage to combine both good clarity and reverberation through a combination of absorbing stage elements and a large internal volume. Both features must be present. Internal volume is expensive, and adding absorption to the stage can be politically difficult, so many small halls lack these features. But if there is enough stage and audience absorption to give good clarity throughout the hall, it can be relatively simple and inexpensive to increase the late reverberation through a modern electro-acoustic system. Properly designed—and this is not always the case—these systems increase the late reverberation time without reducing clarity, and transparently add substantial beauty to the sound. In at least three major opera houses and spaces of all shapes and sizes these systems have been operating for more than twenty years with excellent reviews from the critics and the public.

Postlude

This paper is not as controversial as it might seem. The model of hearing we present is similar to the latest work on the subject, particularly the model proposed by Torsten Dau at the Danish Technical University. The sections on stream formation and its effects on sound are found in standard literature such as "Auditory Scene Analysis" by Bregman. Comb filters are also not new, having been proposed by Peter Cariani. What may be new is our proposal that separation of sound elements by pitch can precede their analysis for timbre and direction, and that the information necessary for this separation lies in modulations induced by the phases of upper harmonics. We have found the physics needed to make this idea work, made a model of the process in the C language, and shown that when the model operates on binaural recordings of live music it predicts the point in a hall where localization disappears.

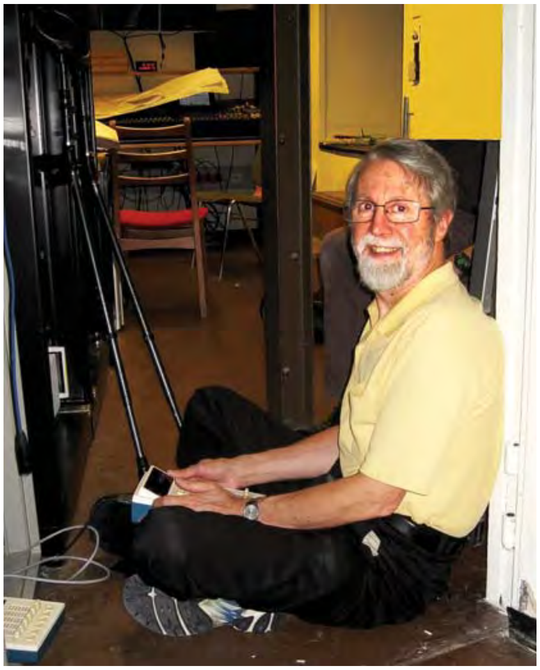
Perhaps the most controversial proposition is that popular acoustical thinking is incorrect in believing that more early lateral reflections are always good, that clarity can be measured by standardized measures such as C80 and C50, and that the strength of reflections and reverberation should be independent of frequency. In our view when the direct sound is weak early reflections from any direction, but especially medial reflections (those from the front, rear, and ceiling), are detrimental to the sound. If clarity is defined by the ability to distinctly hear the notes in a performance, a high value of C80 or C50 often predicts the opposite. The best of the standard measures, IACC80, is insensitive to medial reflections. But at the recent International Conference on Acoustics in Sidney a keynote speech by Leo Beranek and several other papers called into question the reliability and even the relevance of these measures. The field of room acoustics seems open for change. Hopefully the neurology of hearing will play a prominent role in this process. **AT**

Reference

1 With the permission of the Pacifica String Quartet we can hear two examples from a concert in a 1300 seat shoebox hall. The sound in row F is quite different from the sound in row K. The recordings are from the author's eardrums, and are equalized for playback over loudspeakers or headphones equalized to sound identical to loudspeakers. (Most headphones have too bright a sound to reproduce them correctly. Pink noise played though the headphones should sound identical in timbre to the same noise played through a frontal loudspeaker.) Instructions for downloading the audio clip examples are given in the sidebar.

"Binaural Recording of the Pacifica String Quartet in Concert row F"; (http://www.davidgriesinger.com/Acoustics_Today/row_f_excerpt.mp3)

"Binaural Recording of the Pacifica String Quartet in Concert row K"; (http://www.davidgriesinger.com/Acoustics_Today/row_k_excerpt.mp3)



David Griesinger in the Schillertheater Berlin. Photo taken by Albrecht Krieger, Tonmeister of the Deutches Staatsoper.

David Griesinger is fascinated by the relationship between mathematical science and the recording, reproduction, and perception of music. He has worked as a classical music recording engineer all his life, an avocation that encourages a certain skill in listening to sound. He has also been active as a singer in various music groups, including the Boston Camerata. After completing his Ph.D. in physics in 1978 he independently developed one of the first digital reverberation devices, later to become the Lexicon 224. A more than thirty year stint as chief scientist for Lexicon followed, leading to many products, such as the LARES reverberation enhancement system and the Logic7 surround system. He has given lectures and papers on recording and room acoustics around the world, taking time to listen to as many concerts as possible. He now works independently on the mechanisms the ear and brain use to perceive sound, and how these mechanisms are affected by reverberation. He lives in Cambridge Massachusetts, where enjoys his family, concerts, and making HD video recordings of musical performances.

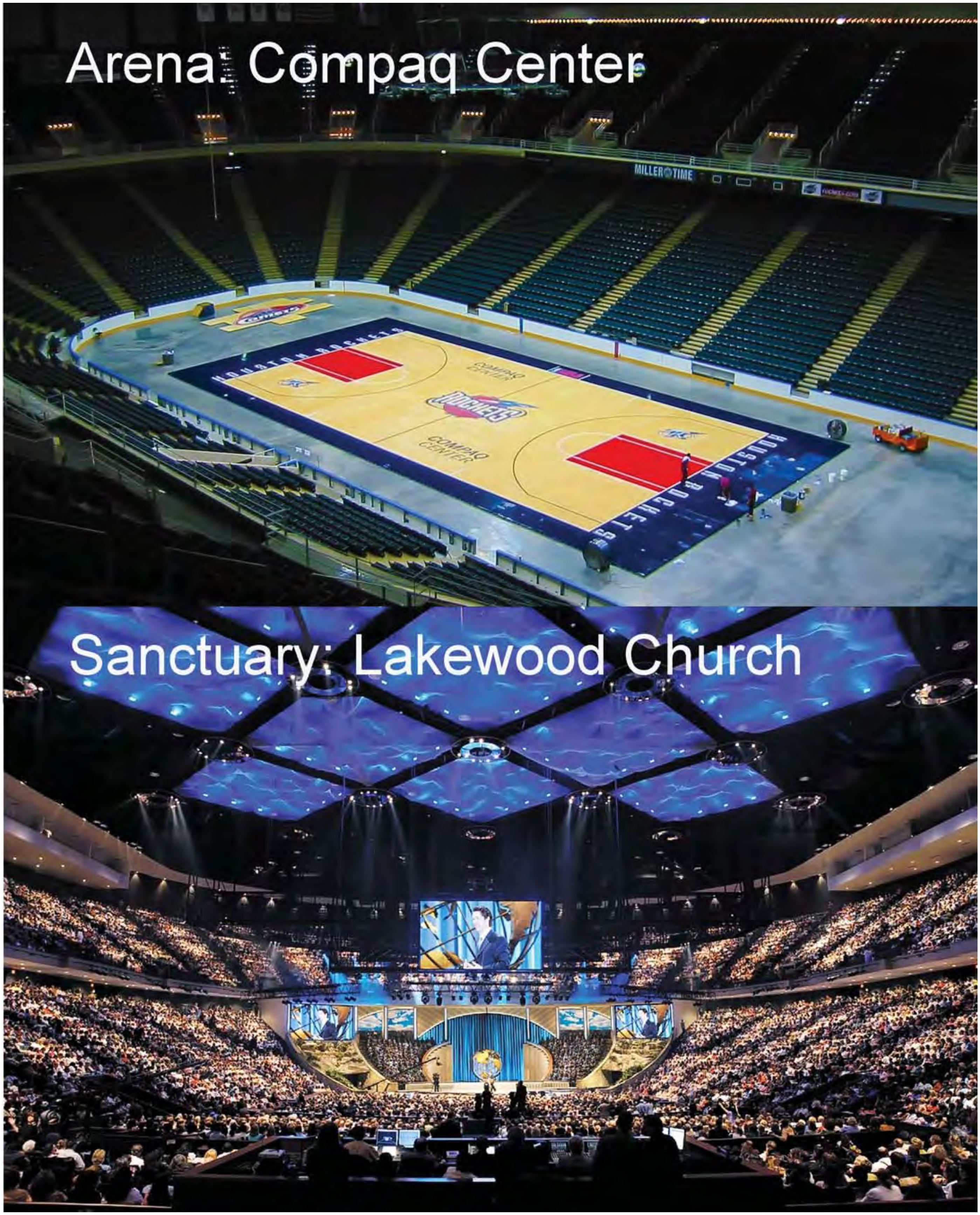


Fig. 1. Compaq Center is now Lakewood Church.