EQUALIZATION IN SOUND REINFORCEMENT: PSYCHOACOUSTICS, METHODS, AND ISSUES

Russell Ovans

Jason Sound Industries, 1709 Welch St., N. Vancouver, BC V7P 3G9
and
School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6
ovans@cs.sfu.ca

SUMMARY

The practise of electroacoustic sound system equalization demands an understanding of psychoacoustics, room acoustics measurement, and subjective room acoustics. In this paper we undertake a review and synthesis of the literature pertaining to the relevant psychoacoustic and room acoustic phenomena, and relate it to a number of issues regarding the current methods of large venue sound system equalization.

SOMMAIRE

L'utilisation de systèmes équalisateurs de son électroaccoustique exige la compréhension de la psycho-accoustique, des mesures de l'accoustique de salle et de l'accoustique subjective de salle. Nous présentons dans cet articles une étude de synthèse de la litérature relative aux phénomènes de la psycho-accoustique et de l'accoustique de salle et relions cet étude à de nombreux problèmes qui se rapportent aux méthodes courantes pour les systèmes équalisateurs de son pour grande salle.

1. INTRODUCTION

Large venue sound reproduction systems consist of a network of signal processing equipment through which an original source signal is routed on its journey to multiple speakers at various locations within a room. As part of this network, equalizers are often used to modify the frequency spectrum of the source signal before it is fed to the speakers, in an effort to compensate for unevenness in the frequency response of both the speakers and the venue. In touring systems, how the audio engineer chooses to do the time-constrained and complex task of adjusting the equalizers is as much a black art as it is a science.

Our main objective in writing this paper is to review and hopefully synthesize much of the research relating to equalization as it pertains to electroacoustic sound reinforcement technology in large venues. While equalization relates to the way very large loudspeaker systems function in any acoustical environment (including outdoors), the discourse is restricted to enclosed spaces. The practise of room equalization requires an understanding of the science of room acoustics and the psychoacoustic considerations of mapping objective, quantifiable measurements to subjective listener preference. In this respect, equalization is closely related to the field of subjective room acoustics.

In this paper we first consider the relevant psychoacoustic (Section 2) and room acoustic phenomena (Section 3) before introducing a number of issues regarding the practise of sound system equalization (Section 4). Though our particular perspective is touring systems and audio engineering, it is our belief that the principles discussed are also of interest to acoustical consultants for fixed installations.

2. PSYCHOACOUSTIC PRELIMINARIES

Psychoacoustics is the specific branch of psychophysics concerned with the relationship between the objective, physical, and quantifiable properties of sound stimuli in the environment and the subjective, psychological, and qualitative responses they evoke [Rasch82a]. There are two important psychoacoustic issues with respect to equalization: the frequency response and critical bandwidth of the ear.

2.1 Frequency Response of the Ear

Pitch is the perceptual correlate of frequency. However, this correlation is not linear: as the number of cycles per second increases linearly, our perceived sense of pitch increases only logarithmically. Alternatively, as our sense of pitch increases linearly, the frequency increases exponentially. For example, the doubling of frequency with every octave represents an exponential growth in frequency as our pitch impression grows linearly.

The range of human hearing is well known to be 20 - 20,000 Hz. However, perception is not equally sensitive at all frequencies; i.e., the ear does not exhibit flat frequency response. Fletcher and Munson's famous curves of equal loudness [Fletcher33] illustrate quite clearly that the ear's sensitivity to loudness is frequency dependent. The main characteristic of the F-M curves is decreased sensitivity at low and high frequencies, but as intensity increases, sensitivity flattens out. At any level, maximum sensitivity occurs at about 3 kHz, corresponding to the resonant frequency of the ear canal [Houtsma87].

The F-M curves were determined using a small set of pure tones in an anechoic space with the sound source directly in front of the test subjects. The listeners had one ear blocked with cotton balls soaked in Vaseline. The F-M results are therefore a measure of the monaural perception of pure tones in a free field with an on-axis sound source.

However, none of the succeeding studies are in agreement with the F-M curves [Holman78]. In particular, the ISO adopted curves for free field listening are parallel at all levels above 400 Hz. So while the ear is not flat at high frequencies, it does exhibit the same response regardless of level. Moreover, the frequency response of the ear depends on the sound field [Holman78] and the position of the sound source [Fletcher53]. According to ISO standard 454, in a diffuse field, the ear is +3 dB more sensitive at 1 kHz, -2 dB at 2.5 kHz, and +4 at 10 kHz. Staffeldt and Rasmussen have shown these numbers to be an approximation of the directional sensitivity of the ear to high frequencies, due to the diffraction caused by the head, torso, and ears. Perceptual sensitivity is a function of distance from the sound source and the room size [Staffeldt82]. The distance from the source changes the diffraction caused by the head and external ear. The size of the room influences the amount of diffusion.

Particularly important is the diffraction due to the pinnae, or outer ear flaps. At its simplest, the pinna acts as a low-pass filter for sounds from behind the head, which provides a cue for distinguishing front from back for high frequency sounds. Research in the 1970s produced convincing evidence that additional localization cues are provided by the reflections of the incident sound off the intricate ridges and depressions of the pinna. These reflections introduce short time delays that are manifest as high-Q notches in the frequency response starting at approximately 6 kHz [Rodgers81]. Because of the geometry of the pinna, as a sound source is raised in elevation the first prominent notch in the frequency response occurs at a higher and higher frequency. Kendall and Martens later asserted that we use these head-related transfer functions as a mechanism for localization on the vertical and front/back planes [Kendall84].

In summary, the frequency response of the ear is dynamic, depending on the listening environment, loudness, and position of the sound source.

2.2 Critical Bandwidth

The basilar membrane – the main sensing mechanism of the ear – is a 35 mm long spiral coil that bulges at a frequency dependent location in response to sound stimuli. The *critical bandwidth* for a given frequency is the smallest band of frequencies around it that will activate the same part of the basilar membrane [Truax78]. Perceptually, the critical bandwidth is the ear's resolution of discrimination; i.e., its resolving power for *simultaneous* tones.

Plots of the size of critical bandwidth as a function of centre frequency indicate that the bandwidths lie between 1/3-octave and 1/6-octave for frequencies above 400 Hz [Houtsma87]. Below 400 Hz the bandwidth is more or less constant at a rather staggering 100 Hz. 24 critical bands traverse the length of the cochlea and therefore define the range of hearing. However, critical bands are different than 1/3-octave analyzers in that "the set of critical band filters is continuous; that is, no matter where you might choose to set the signal generator dial, there is a critical band centered on that frequency" [Everest89, p. 32].

An understanding of critical bandwidth is important to the practise of equalization as it is often (erroneously) cited as a psychoacoustic basis for choosing a particular measurement resolution. Critical bandwidths are more directly relevant to theories of consonance and dissonance (i.e., the subjective agreeability or disagreeability of simultaneous sounds). Two simultaneous pure tones within a critical bandwidth of each other, but not of the exact same frequency, are perceived as dissonant. The two tones result in beats if close together, roughness if further apart, until finally breaking into separate distinguishable tones once they differ by the limit of frequency discrimination. Consonance results only once the tones cross the critical difference and henceforth differ by at least the critical bandwidth. Sounds with spectral content that cross critical bandwidths are perceived as louder than sounds that do not, even if the two sounds have equal rectangular area of sound intensity (defined by intensity per Hz).

3. MEASUREMENT AND EVALUATION

There is nothing quite as upsetting as viewing one's first attempt at measuring the 'frequency response' of a room [Everest89, p. 205].

The three core objective parameters of sound quality for a room are reverberation time, frequency response, and the impulse response. Reverberation time and frequency response are both derivable from the impulse response. In order to understand the impulse response, the room first must be understood as a linear system.

3.1 Linear Systems

A linear system is a mathematical abstraction used to describe any system where the relationship between the output and input is governed by a linear differential equation with constant coefficients:

$$x(t) \longrightarrow \boxed{\text{Linear System}} \longrightarrow y(t)$$

where $x(t)$ and $y(t)$ are time domain representations of signals; for our purposes, these functions represent time-varying sound pressures.

The other predominant way to represent a signal is the frequency domain, whereby a signal is described by the presence of energy at certain frequencies. The two domains are duals: equivalent information is contained in each. Transforms are mathematical tools that enable the movement from one domain to the other.

The Fourier transform is a method of converting between the time and frequency domains. Named for the French mathematician, it is based on his famous theory which states that any periodic time-varying signal can be expressed as the sum of an infinite series of sine and cosine terms each with a specific amplitude and phase. If $x(t)$ is a signal, its Fourier transform $X(f)$ is a function that maps frequency onto a complex number $A + Bi$. The amplitude of the signal content at frequency f is the magnitude of the complex number (i.e., the square root of $A^2 + B^2$) whereas the phase (i.e., its relative alignment) is given by the argument ($\theta = \operatorname{atan}(B/A)$). The frequency domain therefore consists of both an amplitude response and phase response. By convention, the term *frequency response* refers to the amplitude response only.

In practise, one is restricted to discrete time signals obtained with a particular sampling interval. The fast Fourier transform (FFT) is the name for a class of algorithms that quickly compute the Fourier transform of a discrete time signal.

Every linear system is completely described by either its impulse response, or its amplitude and phase response. The impulse response, $h(t)$, is a system's output to the delta function. The delta function, $\delta(t)$, is defined by:

$$\delta(0) = 1$$
$$t \neq 0 \Rightarrow \delta(t) = 0$$

The output $y(t)$ of a linear system to an arbitrary input $x(t)$ is the convolution (*) of the input and the impulse response:

$$y(t) = x(t) * h(t)$$

Convolution works as follows. Each $x(t_i)$ can be thought of as a scaled and delayed version of the delta function. Each $x(t_i)$ therefore produces a scaled, delayed version of $h(t)$. The output $y(t_i)$ is the sum of the scaled, delayed versions of $h(t)$ as generated by each $x(t_j)$, $j \leq i$.

Convolution in the time domain is equivalent to multiplication in the frequency domain:

$$y(t) = x(t) * h(t)$$

$$Y(f) = X(f)H(f)$$

where $X(f)$, $Y(f)$, and $H(f)$ represent $x(t)$, $y(t)$, and $h(t)$ respectively in the frequency domain. The transfer function of a system is its ratio of output to input expressed in either the time domain or the frequency domain. The impulse response is the time domain representation of the transfer function. The frequency domain representation is given by:

$$H(f) = \frac{Y(f)}{X(f)}$$

i.e., the ratio of output to input.

If $h(t) = \delta(t)$, then $y(t) = x(t)$ and $H(f) = 1$, for all f. That is to say, a system with perfectly flat frequency response and no gain or attenuation has an impulse response equal to the delta function.

3.2 Interpreting the Impulse Response of a Room

A room's effect on sound can be modeled as a linear system. Since the impulse response completely defines a linear system, all characteristics of interest are derivable from it, including in the case of a room, reverberation time and frequency response. The traditional method of obtaining a room impulse response is to excite the room with an impulse and to record the decaying sound pressure. A recent paper by

Norcross and Bradley compares four competing approaches to obtaining the impulse response of a room, each of which is shown to produce similar results [Norcross94].

The true impulse response h of a room is an oscillating signal of amplitude vs. time. In acoustics literature, the impulse response is often instead illustrated as the squared impulse response p^2, which is a plot of the square of the amplitude vs. time. The energy-time curve (ETC) is another non-negative real-valued alternative to the impulse response. The ETC and its calculation are described in [Duncan88].

Given an impulse response, the frequency and phase response is contained within its Fourier transform. The reverberation time (RT_{60}) is the length of time it takes the impulse response to attenuate 60 dB. $RT_{60}(\lambda)$ is the reverberation time when the room is excited not with an impulse (which has equal energy per frequency), but with a pure tone of wavelength λ.

In applying linear systems theory to room acoustics, an important consideration is that every pair of source-receiver locations defines a different transfer function. The room as a whole does not possess a single impulse response, but rather defines one for each pair of possible locations. RT_{60} is therefore a function of not only frequency, but also location. "When reverberation time for a given frequency is reported, it is usually the average of multiple observations of each of several positions in the room. This is the pragmatic way of admitting that the reverberatory conditions differ from place to place in the room" [Everest89, p. 207]. Only in a perfectly diffuse sound field is RT_{60} the same at each frequency and location. However, "one still talks of 'a concert hall with RT_{60} of 1.8 s,' as if true for all frequencies and true for all source-receiver combinations in the hall" [Barman93]. The sensitivity to position can be striking: changes in distance as little as 10 cm of either the source or receiver can result in statistically significant changes in measured values of the early decay time (EDT, the first 10 dB of decay) [Bradley89]. Another controversial and inconclusive study [Barman93] reported differences of over 0.5 s at different measurement locations, and 0.6 s at different frequencies for the same location in a large room.

Interpretation of the impulse response and reverberation time is most closely associated with the field of subjective room acoustics. Subjective room acoustics is the psychoacoustic study of perception in enclosed spaces with the goal of determining the important quantitative variables in the design of concert halls and auditoria. According to Rasch and Plomp in the introduction to their excellent survey of the field, subjective room acoustics is the study of the perceptual effects of the indirect sound field, which is responsible for what is loosely called the *acoustics* of a room or hall [Rasch82b]. The indirect sound field is comprised of sound that arrives after one or more reflections. Indirect sound is further classified as either early reflections if it arrives within 50 ms of the direct sound, or reverberant sound otherwise. Depending on the context, early reflections are often counted as part of the direct sound. The indirect sound has three effects [Rasch82b]:

1. it adds sound energy resulting in a perceived increase in loudness;

2. it arrives later than the direct sound, thus reducing definition as it masks the preceding direct sound; and,

3. it arrives from other directions than the direct sound, resulting in a perceived spaciousness.

Although reverberation time is now considered inadequate as a single objective descriptor of room quality [Bradley90], it was once fashionable to consider the question, what is the optimal reverberation time? A thorough theoretical examination of this question is given in [Mankovsky71] where a number of plots of optimal reverberation time versus room size are shown to be inconclusive. Certainly a high reverberation time is a problem for speech intelligibility as it means the masking of new information by old. But higher reverberation times (1.5 to 2.1 s) are acceptable and even desirable for music, particularly romantic classical music [Rasch82b]. The lesson to note is that flat frequency response is not optimal for the reproduction of music.

In current thinking, the ratio of direct to indirect sound is considered more important than reverberation time in predicting the quality of a room's acoustics, particularly with respect to speech intelligibility. As such a number of measures of room quality have been proposed that are based on the ratio of two integrals of the impulse response (see for example [Bradley90, Davis87]). These ratios of integrals usually differ only in how much of the early reflections count towards the direct sound. Citing the Haas effect [Haas72], many descriptors count early reflections that arrive within 50-80 ms as part of the direct sound. These newer ratios of direct to indirect sound are important, Bradley and Halliwell claim, because they relate well to the subjective assessments of the acoustical characteristics of halls [Bradley89].

However, methods of prediction based entirely on the time domain (i.e., RT_{60} and direct/indirect ratios) suffer two serious shortcomings due to the following limitations of the impulse response as a representation device.

1. It provides no clues as to the directionality of the reflections. The lack of information concerning the directivity of the reflections is critical since "after it was established that early reflections are subjectively important, the direction of arrival of these reflections was next found to be important"

[Bradley90, p. 17]. In particular, early lateral energy (first reflections from the side walls) is considered of fundamental import [Barron71, Schroeder84]. In summarizing the various ratios of energy as a predictor of room quality, Rasch and Plomp note the importance of sound coming from the sides and from the rear later than 40 ms and earlier than 80 ms after the direct sound [Rasch82b]. Because it arrives earlier than 80 ms, it functions as direct sound and improves clarity. Since it arrives from the sides and back after 40 ms, it increases the sense of spaciousness. Since auditorium design is generally a trade-off between clarity (for speech) and spaciousness (for music), they conclude that these reflections are potentially very important. In an attempt to measure directional characteristics of reverberation, Abdou and Guy developed a PC-based measurement system that employs six microphones arranged in cartesian coordinates [Abdou93]. Their system captures the temporal arrival, direction, and magnitude of reflections and plots this information as a series of intensity vectors in time.

2. It provides no immediate clue as to the frequency response. While it is true that frequency domain information is contained within the impulse response, it is not obvious what it is simply from inspection. In other words, the time and frequency domain representations share the same information, but their respective representations are more amenable to the extraction of different information. For example, Toole has criticized the impulse response because it is inferior to the frequency domain for the identification of audible resonances [Toole86a]. Given that the goal of the direct/indirect ratios is to move towards understanding what a desirable transfer function for a room is, one wonders why work in subjective room acoustics seems universally restricted to the time domain. Since the amplitude response is basically a picture of the relative RT_{60} along the frequency axis, it seems strange to dismiss it as a tool. Conversely, audio engineers working with electroacoustic sound reinforcement systems operate exclusively with frequency domain representations.

There are, however, many valuable guidelines to be learned here from the work in subjective room acoustics. In particular, the importance of directivity of reverberation. Moreover, it would be instructive and interesting to consider both the impulse *and* frequency response measurements of halls judged to be excellent.

4. EQUALIZATION

Equalization is the purposeful alteration of a signal to add and/or remove spectral content. From an engineering perspective, equalization is the deconvolution or inversion of the transfer function of a linear system. This is based on the assumption that all artifacts of the intervening linear system are unwanted.

With sound reinforcement systems the common practise is to alter the frequency spectrum of a signal using a 1/3-octave equalizer, which is a collection of 30 independent bandpass filters that each can boost or cut the signal by approximately 12-dB at their centre frequency. This compensation, applied just before the amplifiers in the audio chain, is to correct for aberrations in the response due to the interaction of the loudspeakers and room. The goal of this compensation is twofold [Davis87]:

1. to ensure a specified tonal response at each listener's ears; and,

2. to maximize overall acoustic gain by reducing peaks in the frequency response that can cause the system to enter a feedback loop.

Implicit in point 1 is that the specified tonal response will result in improved sound quality. To that end, equalization of the frequency (amplitude) response is considered the single most important method of improving the listener preference rating of a loudspeaker [Fortier94].

ISO standard 2969 describes a recommended method for equalization: excite the room with pink noise, measure the frequency response using a 1/3-octave real-time analyzer (RTA), and adjust the equalizer until a certain curve is realized on the RTA. The frequency response of a venue as viewed on a 1/3-octave RTA is called the house curve. Pink noise is used because it has equal energy per octave. White noise, because it has equal energy per Hz, exhibits a +3-dB/octave rise in energy with increasing frequency and is therefore less suitable when using constant percentage bandwidth filters such as those used in 1/3-octave analyzers.

This ISO standard raises many issues and questions:

1. What is the ideal house curve?
2. Do you equalize based on measurement of the direct, indirect, or total sound?
3. Is a single measurement point adequate?
4. Should the audience be present?
5. What about time domain equalization?
6. Is 1/3-octave resolution enough?

In the remainder of this section we consider each of these questions in turn.

4.1 The Ideal House Curve

The immediate question arises: what is the ideal house curve? Toole has presented convincing evidence that in anechoic conditions, listeners prefer loudspeakers with the smoothest and flattest frequency response, both on- and off-axis [Toole86b]. Conversely, the Athena project has suggested that in a typical small listening room flat frequency response is **not** the optimal transfer function [Fortier94], but project participants have not revealed what they believe it to be.

Given this inconsistent state of affairs, let us first consider what equalization should attempt to correct. Bucklein has examined the effect a nonuniform frequency response has on speech intelligibility over telephone lines [Bucklein81]. The result of his study, which also held for music and white noise, was that peaks in the transfer function are clearly more disturbing than corresponding valleys. Satisfactory intelligibility requires that narrow peaks must be avoided, while several small valleys, even if these are deep, are tolerable. Test subjects perceived no difference in the source material if the transfer function contained a single 5-dB valley an octave wide. The narrower a valley becomes, the greater its depth must be to remain audible; e.g., a 20-dB dip with bandwidth $\Delta f / f = 0.2$ was judged inaudible at all frequencies measured (NB: $\Delta f / f = 0.23$ for 1/3-octave). The subjective judgement is also worth note: the listeners reported that the audible valleys do not alter the sound quality as much as equally large peaks, which can appear "very unpleasant." If peaks are unavoidable, two narrow peaks are better than one wide one, and the farther apart, the better. Note the consistency with critical bandwidth theory, which predicts that wide peaks that cross critical bandwidths will be perceived as louder than narrow peaks that do not. A number of widely spaced peaks is better than a single wide peak in terms of intelligibility (and corresponding tonal colouration).

Current guidance – as espoused in for example [Davis87] – is that one should measure the house curve with a flat response free-field microphone placed about 30 m from the source. The equalizer should be adjusted so that the house curve is flat up until about 1 kHz where a roll-off down to -10 dB at 10 kHz should begin. What explanation is there for this high-frequency attenuation? Papers by Schulein [Schulein75] and Staffeldt and Rasmussen [Staffeldt82] address this question. Taken together, these two papers are crucial in understanding the psychoacoustic considerations of equalization.

Schulein considers the question of the high-frequency roll-off: why is it that a flat house curve, obtained by exciting a sound reinforcement system with pink noise and viewing on a 1/3-octave analyzer, sounds too bright? Through an ingenious experiment, Schulein deduced two causes: the increased sensitivity of the human auditory system to high-frequency diffuse sound as opposed to near-field frontal sound; and, the roll-off in diffuse-field sensitivity versus free-field sensitivity in commercially available measurement microphones. "Due to the polar characteristics of the human listener, a lower sound pressure level is required for equal loudness at high frequencies for a diffuse sound field than for a frontal sound field" [Schulein75, p. B-47]. He cites the shape of the head as the culprit, and suggests the design of microphones that mimic the resulting directional pattern. Microphones embedded in dummy heads would seem a more expedient alternative.

Schulein's view is reinforced by the work of Staffeldt and Rasmussen [Staffeldt82]. The important points are as follows. If a human equalizes two loudspeakers such that they sound equally loud at all frequencies, and one of the loudspeakers is in the distance such that it produces a reverberant field, and the other is within the critical distance such that it mimics a free field – the distant loudspeaker producing a diffuse field will have its high frequencies attenuated. Equivalently, if you replace the human with a microphone and do equalization such that the measured frequency response of both loudspeakers is flat, the diffuse field loudspeaker will sound brighter. However, if you embed that microphone inside the ear of a dummy head and repeat the process, the perceived brightness disappears. It is not adequate to use an omnidirectional microphone with flat free field sensitivity and flat diffuse field sensitivity; the dummy head is necessary. In fact it is psychoacoustically invalid to use an omnidirectional microphone: as noted before, the human ear is not uniform in directivity at high frequencies. For example, the ear is +10-dB more sensitive at 6400 Hz to sound 90-degrees off-axis than it is to sound on-axis [Fletcher53]. In a diffuse field, where sound is entering the ear from all directions, this sensitivity is stimulated.

Moreover, high frequency response is dependent on distance to the source (regardless of whether the listener is inside or outside the critical distance): an equalized loudspeaker in anechoic conditions will sound different with distance. This is because the free-field response of the external ear depends on the distance between the head and the loudspeaker:

> *...It is concluded that the high-frequency attenuation necessary for a distant loudspeaker when compared with a nearby loudspeaker is largely determined by* **[BOTH]** *the free-field and diffuse-field diffraction phenomena at the head and the external ear* [Staffeldt82, p. 642].

As a caveat, Staffeldt and Rasmussen warn that these results may not be generalizable to large venues.

4.2 Direct vs. Indirect Sound

Modern sound reinforcement systems for large spaces are almost always comprised of multiple loudspeakers. The use of multiple source positions impacts the perceived reverberation of an enclosed space.

The critical distance is the point at which the intensity of the direct field is the same as that of the reverberant field. Beyond the critical distance, the ratio of direct to indirect sound steadily degrades. Ideally then every seat in the audience should be within the critical distance. Unfortunately, for all venues of any significant size, most of the audience is beyond the critical distance. That is, most people are listening to the reverberant sound field more than the direct field. For example, an omnidirectional sound source in a large concert hall (volume = 27,000 m^3, RT_{60} = 2.2 s) has a critical distance of 11 m [Plomp73]. Since intelligibility and clarity are proportional to the ratio of direct to indirect sound [Rasch82b], sound quality can degrade significantly beyond the critical distance.

Employing multiple loudspeakers is not the solution. In fact, multiple loudspeakers is part of the problem. Every loudspeaker that is added contributes to the indirect field and therefore degrades the direct/indirect ratio. The optimal direct/indirect ratio is obtained with a single radiating point.

O'Keefe has looked at the problem of critical distance and multiple loudspeakers in very large reverberant spaces. "...The fundamental dilemma associated with very large rooms: increasing the number of speakers means that some people will be exposed to better direct and early sound. For people located elsewhere in the room these same loudspeakers will introduce detrimental late sound" [O'Keefe94, p. 71]. For the Galleria in Toronto, a 90,000 m^3 space, O'Keefe separately calculated and measured the direct, early, and late sound and plotted their intensity as a function of distance from sound source. The direct and early sound levels decayed linearly with similar slope; the late sound was virtually constant. The point at which the direct and reverberant lines cross is of course the critical distance, which he found to be 7 m. He empirically noted that beyond 7 m speech intelligibility decreased significantly. This was for a single loudspeaker: "the important difference between a single loudspeaker system and a distributed system with several loudspeakers is that the distant loudspeakers generate sound that a listener will interpret as late or detrimental" [O'Keefe94, p. 72]. As loudspeakers are added, the critical distance drops. With 16 loudspeakers, their location and spacing became insignificant for listeners more than 2 or 3 m from the nearest speaker, leading O'Keefe to conclude that no matter where one stood in the room, there must be a loudspeaker within 3 m. A grim conclusion to say the least.

This begs the question, why do concert sound systems rely on massive arrays of speakers? The conventional wisdom is that the best way to battle the critical distance problem is by increasing the intensity of the direct sound.[1] A large semi-circular array of loudspeakers is believed to deliver a higher ratio of direct to indirect sound to all portions of the audience, though this is theoretically a dubious claim.

In large reverberant spaces where most listening locations are subjected to a poor direct/indirect ratio, a question to consider is what do you equalize: the direct, indirect, or total sound? Meyer claims that "it is known that the ear generally perceives early reflections as the 'frequency response' of the space" [Meyer, p. 3]. Meyer's technique is to use correlation with the excitation source to reject reverberation in his measurements. Truncation of the impulse response also provides a means of considering just the direct sound [Genereux90]. But does it make psychoacoustic sense to reject the predominant sound field?

Cabot [Cabot88] has noted that the pink noise RTA technique results in a measurement of the steady-state room response, or the integral of the direct and indirect sound. Consistent with [Staffeldt82], Cabot finds that "... a flat sounding system will usually not be flat in the direct field response or in the steady state response," which leads him to conclude that "it is therefore as incorrect to equalize the steady state response as it is to equalize the direct sound" [Cabot88, p. 392]. Missing from his analysis, unfortunately, is an explanation due to the high-frequency directional sensitivity of the ear.

Toole and Olive found that a loudspeaker is judged favourably when on- and off-axis response are both flat, whereby the direct and indirect sound match [Toole88]. While they did not make the connection, the Haas effect [Haas72] is likely responsible. In the case of matching early reflections, the Haas effect holds and the direct sound is reinforced [Rodgers81]. If, however, the frequency response of the early reflections differs significantly from that of the direct sound, the Haas effect is defeated and the early reflections will be perceived as annoyingly audible, helping to further degrade the direct/indirect ratio.

This suggests that one might consider equalizing the indirect sound such that it matches the direct sound. (Unfortunately, this is complicated by the fact that you can't change one without affecting the other.) Sliding the FFT time window to the latter portion of the impulse response permits the measurement of the frequency response of the indirect sound field. Alternatively, assuming a flat free field response (a reasonable assumption for most loudspeakers) one could improve the likelihood of matching direct and indirect frequency response through the addition of off-axis full-range loudspeakers. This is the concept behind the design of Bose loudspeakers, and the by-product of any semi-circular array configuration.

4.3 Single vs. Multiple Measurement Points

As noted above, the transfer function of a room differs from location to location. Plomp and Steeneken attempted to cod-

1. This is only part of the answer. A significant factor is that the artist thinks it looks cool. Another reason is that it has to sound like a rock concert regardless of where you are sitting in the audience. For large, highly reverberant spaces, there is only one way to ensure that: **volume**.

ify the amount of fluctuation one can expect in a reverberant space [Plomp73]. In their paper, they show that location dependence is caused by the variability in the amplitudes and phases of the individual harmonics of a complex tone in a diffuse field. The variabilities in amplitude (SPL) as a function of location are derived theoretically to have a standard deviation of 5.57 dB for pure tones in a diffuse field; phase differences are random (0 to 2π) and found to be negligible for complex tones with fundamental frequencies above 100 Hz [Plomp73]. Since timbre is correlated to the relative amplitude of the harmonics of a complex tone, timbre differs from location to location as a function of this variance in SPL. Their empirical study supports this theoretical variation.

Worse, the problem of location dependent frequency response variation can actually be exacerbated by equalization. Elliott and Nelson have empirically shown that optimizing for a single location within a small room is detrimental to all other points within the room [Elliott89]. This phenomenon may or may not generalize to the case of large enclosures, but the anecdotal evidence suggests that it does.The burden of averaging multiple measurement points seems to be the answer. For touring systems, however, there is generally not enough time nor the proper equipment to undertake multiple simultaneous measurements. Moreover, the complexity of adjusting multiple equalizers while simultaneously considering multiple room response curves would quickly result in cognitive overload for the audio engineer. It is for these reasons that the best seat is usually the one next to the front-of-house mixing console.

4.4 Audience Presence

Ideally, the audience should be present before a room is equalized. The audience has two significant effects on a room. First, the audience increases the absorption characteristics and thus affects the reverberation time and the frequency response of the room. Second, the audience increases the temperature of the air, creating temperature gradients that change room mode interactions and thus affect the resulting frequency response.

This is not to suggest that equalization of an empty hall is pointless. The presence or absence of an audience does not greatly affect resonances and echoes that are a function of the dimensions of the hall. Some have even suggested that the audience is not truly significant [Beranek62]. We will only add that equalization without an audience present is better than no equalization at all.

4.5 Time Domain Equalization

Time domain equalization is equalization specified by the impulse response of an arbitrary filter, providing user control over both the amplitude and phase response. Traditional equalizers affect both the amplitude and phase response, but provide user control over only the amplitude response. The digital delay is a degenerate case of time domain equalization consisting of a delta function capable of being offset in time.

The most obvious shortcoming of frequency (amplitude) equalization is that it is not a complete substitute for time domain equalization. "Equalization in the frequency domain effectively only equalizes the minimum phase part of the response due to the presence of all-pass phase components in the very complex room response" [Fortier94, p. 60]. A thorough and theoretical treatment of this phenomenon is Neely and Allen's seminal paper on the invertibility of room impulse responses [Neely79]. [Elliott89] is one of a class of papers dealing with the use of time domain room equalization in the form of adaptive digital filters. An adaptive digital filter is a digital filter that changes in response to an error signal, either continuously, or is "programmed" once from measured data. A series of automatic equalization schemes based on adaptive digital filters have appeared in the literature supporting optimization at a single point with a single filter [Genereux90], multiple points with a single filter [Elliott89], and multiple points with multiple filters [Munshi92]. Equalizing multiple loudspeakers with different filters is the only approach capable of completely inverting the impulse response of a room [Munshi92].

In multi-loudspeaker systems, nulls will occur in the polar pattern as a result of waveform interference. While dips in the amplitude response are correctable in the frequency domain, nulls are correctable only in the time domain [Reams94]. As an example, Davis and Davis cite the comb filter effects produced by misaligned speakers, which cannot be detected with a 1/3-octave analyzer or corrected with a 1/3-octave equalizer [Davis87]. While the audibility of these nulls is questionable, these comb filters can be mistaken for the high-frequency notches used as localization cues [Rodgers81]. The problem is easily solved by aligning the loudspeaker wavefronts via the introduction of a delay.

However, the general consensus is that phase equalization is a distant second to amplitude equalization in importance (e.g., [Toole86a]). Equalization in the time domain improves phase effects that many listeners are insensitive to [Fortier94].

4.6 Measurement Resolution

Many have cited the critical bandwidth of the ear as support for the use of 1/3-octave analyzers and equalizers (e.g., [Schulein75]). If you consider the definition of critical bandwidth put forward by Davis and Davis, it seems particularly apt: the bandwidth within which the human ear cannot detect spectrum shape when listening to complex sounds [Davis87]. In other words, correcting for anomalies in the spectrum at a resolution finer than 1/3-octave is pointless. This is a reasonable first approximation, but is unfortunately incorrect. Critical bandwidths define regions of dissonance, within which it is still possible to detect spectrum shape.

Meyer suggests that high-resolution DFT analysis techniques are by far preferable to 1/3-octave [Meyer84]. Toole and Olive concur: "with any measurement it is clearly important that there be adequate frequency resolution to reveal the presence of high-Q resonances.... The popular 1/3-octave measurements are useful only to reveal gross features in the frequency domain" [Toole88, p. 140].

5. CONCLUSION

In this paper we have looked at the practical aspects and psychoacoustic considerations of room measurement and equalization. Hopefully some questions have been answered, but many ambiguities remain. If this survey is to function as a springboard for further research, one might consider the following questions as a guide.

1. What is the most appropriate method for measuring the frequency response of a sound reinforcement system? Some research has suggested that microphones embedded in dummy heads provide a more reliable measure of perceived frequency response, and that high resolution DFT analysis is preferred to a 1/3-octave RTA. But, how many measurements at how many locations? How should multiple measurements be averaged? Should the source material be pink noise or impulses? Pink noise techniques implicitly measure the steady-state total sound energy. Impulses permit the separation of direct and indirect sound. Which is best?

2. It seems that the choice of target transfer function is dependent on the measurement technique. Many researchers have attempted to explain away these inconsistencies as a function of directional reverberation in rooms and the directional sensitivity of both ears and microphones. Given an understanding of these interactions, is there an ideal transfer function for which equalization should strive? We have pointed out the necessity to eliminate peaks in the frequency response for two reasons: to increase overall gain without causing feedback, and to avoid tonal colouration. But how flat is too flat?

ACKNOWLEDGEMENTS

The author wishes to thank Barry Truax, Jeff Berryman, Jeff Lilly, and the anonymous referees for their helpful insight. The financial assistance of the National Research Council of Canada and the Science Council of British Columbia is gratefully acknowledged.

REFERENCES

[Abdou93] Abdou, A. and R. W. Guy. "A PC Based Measurement System for Obtaining Spatial Information and Objective Room-Acoustic Indicators." *Canadian Acoustics,* 21 (1), pp. 9-14.

[Barman93] Barman, M., A. Gambino, R. Ramakrishnan, and J. C. Swallow. "Reverberation Time?" *Canadian Acoustics*, 21 (3), pp. 79-80.

[Barron71] Barron, M. "The Subjective Effects of First Reflections in Concert Halls – The Need for Lateral Reflections." *J. Sound and Vibration,* 15, pp. 475-494.

[Beranek62] Beranek, Leo L. *Music, Acoustics and Architecture.* John Wiley & Sons, 1962.

[Bradley89] Bradley, J.S. and R. E. Halliwell. "Making Auditorium Acoustics More Quantitative." *Sound and Vibration*, 23, pp. 16-23.

[Bradley90] Bradley, J. S. "The Evolution of Newer Auditorium Acoustics Measures." *Canadian Acoustics,* 18 (4), pp. 13-23.

[Bucklein81] Bucklein, R. "The Audibility of Frequency Response Irregularities." *J. Audio Eng. Soc.*, 29 (3), pp. 126-131.

[Cabot88] Cabot, Richard C. "Equalization, Current Practice and New Directions." *The AES 6th International Conference: Sound Reinforcement*, Nashville, 1988, pp. 389-402.

[Davis87] Davis, Don and Carolyn Davis. *Sound System Engineering (Second Edition).* Sams, 1987.

[Duncan88] Duncan, Andrew. "The Analytic Impulse." *J. Audio Eng. Soc.*, 36 (5), pp. 315-327.

[Elliott89] Elliott, S. J. and P. A. Nelson. "Multiple-Point Equalization in a Room Using Adaptive Digital Filters." *J. Audio Eng. Soc*, 37 (11), pp. 899-907.

[Everest89] Everest, F. Alton. *The Master Handbook of Acoustics (Second Edition).* Tab Books, 1989.

[Fletcher33] Fletcher, G. H. and W. A. Munson. "Loudness, Its Definition, Measurement, and Calculation." *J. Acoustic. Soc. Am.*, vol. 5.

[Fletcher53] Fletcher, Harvey. *Speech and Hearing in Communication.* D. Van Nostrand, 1953.

[Fortier94] Fortier, Claude R. and Pierre Cote. "Digital Signal Processing Applied to the Equalization of the Loudspeaker Room Interaction." *Canadian Acoustics*, 22 (4), pp. 59-60.

[Genereux90] Genereux, Ronald P. "Adaptive Loudspeaker Systems: Correcting for the Acoustic Environment." *Presented at the AES 8th International Conference*, Washington, D. C., 1990.

[Haas72] Haas, Helmut. "The Influence of a Single Echo on the Audibility of Speech." *J. Audio Eng. Soc.*, 20 (2), pp. 146-159.

[Holman78] Holman, Tomlinson and Frank Kampmann. "Loudness Compensation: Use and Abuse." *J. Audio Eng. Soc.*, 26 (7/8), pp. 526-536.

[Houtsma87] Houtsma, A. J. M., T. D. Rossing, and W. M. Wagenaars. *Auditory Demonstrations (compact disc notes).* Institute for Perception Research, Eindhoven, The Netherlands, 1987.

[Kendall84] Kendall, Gary S. and William L. Martens. "Simulating the Cues of Spatial Hearing in Natural Environments." *Proceedings of the 1984 International Computer Music Conference*, pp. 111-125.

[Mankovsky71] Mankovsky, V. S. *Acoustics of Studios and Auditoria.* Translated by Gordon Clough. Edited by Christopher Gilford. Focal Press, 1971.

[Meyer84] Meyer, John. "Equalization Using Voice and Music as the Source." *Proceedings of the 76th Convention of the AES,*

New York.
[Meyer] Meyer, John. "Precision Transfer Function Measurements Using Program Material as the Excitation Signal." Meyer Sound Laboratories Technical Report.
[Munshi92] Munshi, Anees S. "Equilizability of Room Acoustics." *1992 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (vol. 2),* pp. 217-220.
[Neely79] Neely, Stephen T., and Jont B. Allen. "Invertibility of a Room Impulse Response." *J. Acoust. Soc. Am.*, 66 (1), pp. 165-169.
[Norcross94] Norcross, S., and J. S. Bradley. "Comparison of Room Impulse Response Measurement Methods." *Canadian Acoustics*, 22 (4), pp. 47-48.
[O'Keefe94] O'Keefe, John. "Room Acoustics Considerations for Loudspeaker Systems in a Large Reverberant Space." *Canadian Acoustics*, 22 (4), 1994, pp. 71-72.
[Plomp73] Plomp, R. and H. J. M. Steeneken. "Place Dependence of Timbre in Reverberant Sound Fields." *Acustica*, 28 (1), pp. 50-59.
[Rasch82a] Rasch, R. A. and R. Plomp. "The Perception of Musical Tones." In D. Deutsch, ed. *The Psychology of Music.* Academic Press, 1982, pp. 1-24.
[Rasch82b] Rasch, R. and R. Plomp. "The Listener and the Acoustic Environment." In D. Deutsch, ed. *The Psychology of Music.* Academic Press, 1982, pp. 135-147.
[Reams94] Reams, Robert W. "Real-Time Loudspeaker/Environment Performance Management." *Presented at the 97th AES Convention*, San Francisco, 1994.
[Rodgers81] Rodgers, C. A. Puddie. "Pinna Transformations and Sound Reproduction." *J. Audio Eng. Soc*, 29 (4), pp. 226-234.
[Schroeder84] Schroeder, M. R. "Progress in Architectural Acoustics and Artificial Reverberation: Concert Hall Acoustics and Number Theory." *J. Audio Eng. Soc.*, 32 (4), pp. 194-203.
[Schulein75] Schulein, Robert B. "In Situ Measurement and Equalization of Sound Reproduction Systems." *J. Audio Eng. Soc.,* 23 (3), pp. 42-50.
[Staffeldt82] Staffeldt, Henrik and Erik Rasmussen. "The Subjectively-Perceived Frequency Response in Small and Medium-Sized Rooms." *SMPTE Journal*, July, 1982, pp. 638-643.
[Toole86a] Toole, Floyd E. "Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 1." *J. Audio Eng. Soc.*, 34 (4), pp. 227-235.
[Toole86b] Toole, Floyd E. "Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 2." *J. Audio Eng. Soc.*, 34 (5), pp. 323-348.
[Toole88] Toole, Floyd E. and Sean Olive. "The Modification of Timbre by Resonances: Perception and Measurement." *J. Audio Eng. Soc.*, 36 (3).
[Truax78] Truax, Barry ed. *Handbook for Acoustic Ecology.* A. R. C. Publications, 1978.