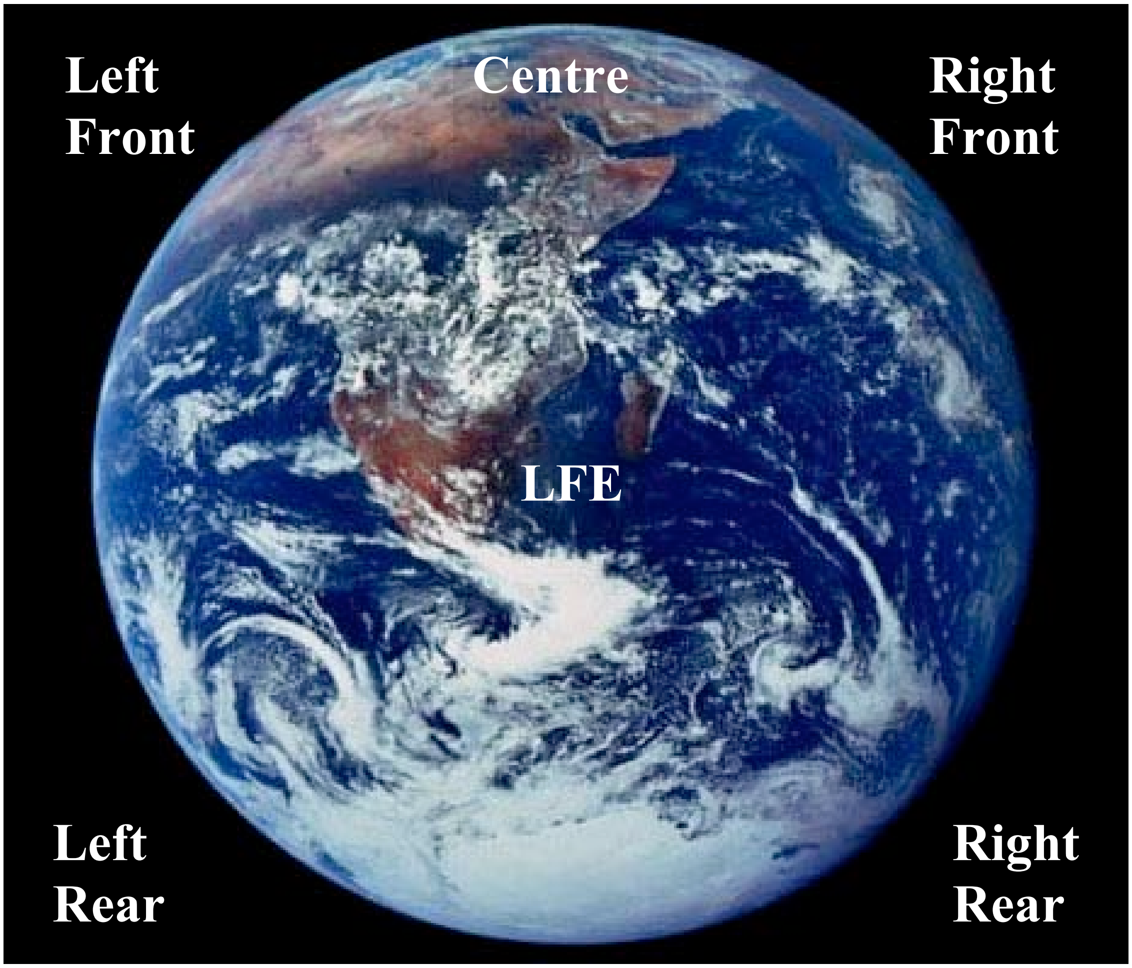
Surround Sound



Today there are various opportunities for up and coming mixers to venture into a surround sound format, especially in their attempts to satisfy the seated audience in a theatre and for those listening and watching on a good home theatre system. The trend these days is that consumers are investing more and more in home theatre systems than traditional stereo systems decreasing the desire to take the time to go a theatre to see the newest Hollywood releases. Just recently film companies are simultaneously releasing films in the theatre, Pay for View and DVD at the same time, causing major film playing theatre chains and distributors to protest on what is sure to be a loss of revenues.

I am sure the initial DVD price of a spanking new release will be close to $30, if not more, until the initial theatre traffic slows down, thus eventually lowering the price of the DVD.

Recently, a joint venture between *Warner Bros.* and the German company *Arvato Mobile* have joined forces in developing a new option for consumers called In2Movies, a P2P delivery system for films. Movies will be available day and date along with physical DVD's, with a price still yet to be determined. Fox studios is following suit also in exploring new distribution methods. In Germany last year almost 12 million movies were downloaded with almost all consumers expressing a desire to pay for downloading services. Will this corrode DVD sales some think? Most executives feel it won't and will add revenues to the $16.3 Billion dollar value. With double digit increases in high bandwidth subscribers, most industry types feel that the market will increase for back catalogues and newer cheaper films smaller independent film companies. If this new marketing strategy proves profitable, look for a major increase in the consumer demand for quality home theatre systems.

Today, post-production professionals have started their own ventures into the post-production business, focusing on providing an excellent product for home theatre systems for the average listening and viewing environment of an average household.
Al Amerod, the winner of 4 Gemini's, recently left Deluxe Studios in Toronto to open his own production facility that will satisfy most of his client demands for a fraction of the price of a major facility.

With the economy looking positive for the next few years to come, I believe that home theatre systems and mp3 cell-phones will dominate the consumer electronic demographic. If we look at Rogers Cable TV, a consumer only needs to purchase a home theatre system and be able to download movies that will allow them all the flexibilities of a DVD except the ownership of a hard product. For the MP3 market, look for a company like Rogers to allow a music lover all the music they want to download for a small monthly fee. If this is to be the case, look for the MP3 format to be abandoned for higher quality audio. Teenagers who represent a large purchasing demographic for entertainment do not have access to credit card or bank accounts, and will be lured to this new model of supplying entertainment as a service instead of a hard product that can be purchased for a monthly fee from their cable TV or cell phone provider.
If this proves to be the case, then the demand for consistent good entertainment will intensify once bandwidth is increased and the industry monetized.
Once this occurs, which I believe will be soon; more medium sized production and music studios will commence employing more people within

the entertainment industry. The executive producers will not be paralyzed by the high cost of expensive technical services and production that a standard film will require to look and sound good in a potential market that might be shifting from major theatres to home systems. The industry is already seeing a trend where clients go to music houses for product and end up having the music house do all the post and mixing on a simple Pro-Tools system.

Will this be the end to quality filmmaking? I believe not!

With newer technology like Final Cut Pro, HD, DAW's and Home theatre, films will be able to look and sound excellent in the comfort of ones home. Let's face the fact that the Ipod-mp3 world has changed the buying decorum of the average music listener.
It is conclusive to state accurately that most of the music listening world cannot tell the difference between an mp3 and a wave file, nor should they. It is the quality of the content that is important, which is why the buying public is mainly interested in purchasing single songs through downloading for a dollar instead of paying $15-$20 for a CD that might only have 2 or 3 good songs.

Look at DVD's. There has been an incredible resurrection of pre- 1995 movies, with a lot of nostalgic libraries being upgraded with better colour correction and audio quality and then released. Do these films look and sound spectacular? Not really, and does it really matter? Most people will always prefer a good story and good music to technical highlights. There has recently been a decline in Box office receipts during the last couple of years, which I believe is due to the lack of quality and content in a film. Test yourself and watch and listen to a movie produced over 30 years ago like "Lawrence Of Arabia." The individual scenes lasted a lot longer than the movies we see today where there is an average edit every 5-10 seconds. The actors had to really act, the cinematographers relied more on visual imagination, and the composers had to score music for longer scenes that had to hold the interest of the audiences.

With this new archetype, I believe we will see a growth in the film and pay-per-view TV industry where the quality of the content will need to remain high, due to the consumers demand for broad latitude in genres of productions and superior substance. With the increase in Internet bandwidth, larger high-resolution screens and surround sound, the demand for quality will be imperative.

The purchasing public is getting accustomed to having their entertainment in an environment that is comfortable for them, be it with ear-buds or viewing a large LCD screen. These days one can simply pause a film when they are hungry for food, passion or need to go to the washroom. I personally don't like the sound of people chattering, eating chips and stepping on my toes in a theatre. Some people say this is ridiculous, but from my experience, watching and listening to "Crash" was more enjoyable on a superb home theatre system than lining up for an hour at the theatre and listening to cell phones go off. I am getting older (29), however, I have two kids who quite agree with me and keep bothering me to spend the $7,000
I would need on a system suitable for their partiality and mine. The big point here is that the home theatre system market is getting cheaper and more affordable for the average consumer with the advantages starting to out weigh the disadvantages to a great extent. Some may argue this point, but the fact is, the trend is going in this direction and looks irreversible.

So what has this got to do with you and surround-sound mixing?

As previously stated, the very expensive, high quality equipment to achieve a great sounding mix has decreased dramatically in the last five years and is still likely to get even cheaper and more versatile.
This allows the budding young mixer to get his foot in the door at the growing number of post facilities of all sizes and show off their creative talents or even start their own business venture.

I know of talented young mixer at small post facility in Toronto and she is presently mixing three shows per week for TV and assists me on the bigger surround film projects. All of this work is done on a Pro Control, Pro Tools HD, Waves plug-ins, Final Cut Pro, 42 LCD monitor and a Tannoy surround system.
With the ability to get exceptional sounding product in this type of working situation, the onus will be on the creative, not the technical, where it should be.

There are simply too many equipment operators with little creative imagination out there. The truth is, if one does not take a productive creative

attitude, they will have to settle on the role of a glorified underpaid operator. One must master their tools before stepping into the landscape of productive, creative high quality recording/editing/mixing.
Creativity! It is a simple law that one cannot focus on creative multi-tasking when their mind is busy trying to figure out operating necessities. It simply doesn't work when one's tools dictate the rate and skill of production!

kd

Recent graduates in the Digital Applications program at Fanshawe College in London Ontario (Music Industry Arts) consistently have demonstrated numerous highly distinct technical skills and creative resources for the demands of the post-production industry. How are they achieving this? Through constant vigilance on where the future is going in the industry and staying on top of the newer innovative trends and meeting the demands of the entertainment consumer.

With educating and challenging oneself, one must at times abandon predictable production techniques and avoid the dictatorship of these so-called experts and their righteous inerrant methodology

Making the transition from stereo to surround sound mixing is to some extent like adapting from mono to stereo. The wider latitudes and options are a welcome format that allows one to be even more creative in their search for elevating audio to higher standards that appeal even more to the average listener. When approaching a 5.1 surround project, I try to vision the sound of the final outcome before even getting started, an approach that worked well mixing in the stereo format. How should I record everything now that I'll have a surround sound palette to fill?
What editing techniques will I employ? What will be the focus in the mix and how can I maximize the quality and environment of the surround sound format for the playback environment?

In this document I am going to explore surround recording and mixing, analysis of conventional methodology and personal ideas on creativity.

This will be a subjective viewpoint based on how I deal with surround sound and it will likely differ with conventional opinions and standardized methods. In my efforts, I hope no one will:

“Release The Status Quo Hounds”

kd

“If one is to maximize the effects of discrete localization in surround sound mixing. One must first investigate how humans perceive the localization of an originating sound source.”

How We Localize Sound

Listening to a sound source and verifying its origin is dictated by the position of the head to the sound source (Direct Path). When the sound arrives to both ears, the time, frequency content and amplitude will be different between the left and right ear. It is important to acknowledge that a sound’s frequency response deteriorates over distance-mostly with high frequencies due to atmospheric conditions.

A sound will reach the ***ipsilateral*** ear (the ear closest to the sound source) prior to reaching the ***contralateral*** ear (the ear farthest from the sound source). The difference between the onset of non-continuous (transient) sounds or phase of more continuous sounds at both ears is known as the interaural time delay **(ITD).**

Similarly, given the separation of the ears by the head, when the wavelengths of a sound are short relative to the size of the head, the head will act as an “acoustical shadow”, attenuating the sound pressure level of the waves reaching the contralateral ear. This difference in level between the waves reaching the ipsilateral and contralateral ears is known as the interaural level difference **(ILD).**

When the sound source lies on the median plane (center), the distance from the sound source to the left and right ear will be the same therefore causing the sound to reach each of the ears at the same time. In addition, the sound pressure level of the sound at both ears will also be the same. As a result, both the ITD and ILD will be zero. As the source moves to the right or left ITD and ILD cues will increase until the source is directly to the right or left of the listener respectively (e.g. ±90 degrees azimuth).

Similarly, when the sound source is directly behind the listener, both ITD and ILD will be zero and as the sound moves to the right or left, ITD and ILD cues will increase until the sound source is directly to the left or right of the listener (on axis).

Separation of ITD (time) and ILD (level) Cues

Although the ***Duplex Theory**** incorporates both ITD and ILD cues, they do not necessarily operate together. ITD's are prevalent primarily for low frequencies, less than approximately 1500Hz, where the wavelengths of the arriving sound are long relative to the diameter of the head and the phase of the sounds reaching the ears can be determined without ambiguity. For wavelengths smaller than the diameter of the head, the difference in distance
is greater than one wavelength, leading to an ambiguous situation, where the difference does not correspond to a unique location. In this situation it is possible to have many frequencies above 1500Hz arriving in phase to the ears. (e.g. The frequency 2Khz can also be in phase with 4Khz, 8Khz and 16Khz for both ears)

For low frequency sounds in which the ITD cues are prevalent and the waves are greater than the diameter of the head, the sound waves experience diffraction whereby, they are not blocked by the head but rather they "bend" around the head to reach the contralateral ear (omnidirectional) . As a result, ILD cues for these low frequency sounds will be very small (although they can at times be as large as 5dB. However, for frequencies greater than approximately 1500Hz, where the wavelengths are smaller than the head, the wavelengths are too small to bend around the head and are therefore blocked by the head (e.g. "shadowed" by the head). As a result, a decrease in the energy of the sound reaching the contralateral ear will result and hence the ILD cue. (See Fig 1)

Audio waves that radiate isotropically (uniformly in all directions) from a source also lose intensity due to the fact that the energy they carry is spread out over an increasingly large area. This is known as the inverse square law.

To conclude, identification of a sound source is determined by the difference in time and phase relationship and amplitude differences. Whew!

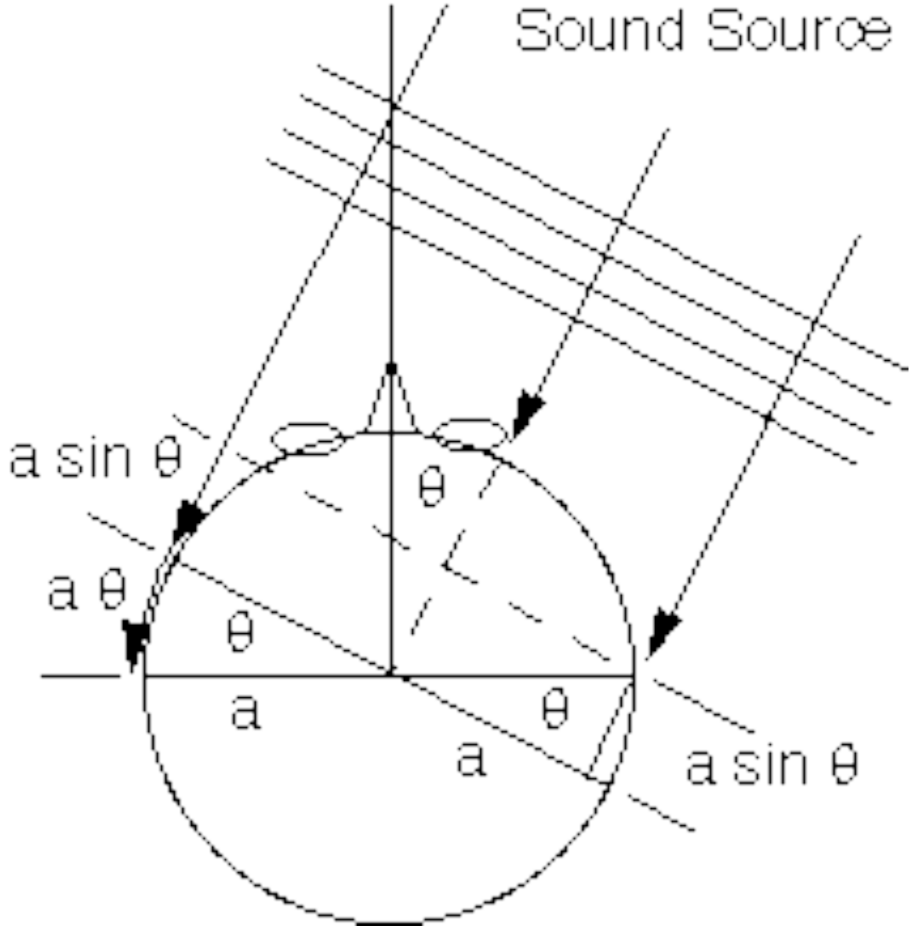


Fig 1: Detecting Sound Source Location (Front-Right)

In Fig 1, we see that early part of the phase of the signal will arrive to the right ear before the left ear (ITD). The level of the signal will be louder in the right ear than the left ear and the mid-high frequency content of the original signal will only arrive to the right ear.

Direct Path (Original) Sound

If you were to suspend two people 10 meters above the ground and 3 meters apart in an open field, you would be able to set up a situation where they could have a conversation with each other with the only audio signal being the direct path route. There would be no floors, ceilings, or walls to reflect the original signal with each listener describing the audio characteristics as being totally dry. If the distance between the two people increase the amplitude and frequency response would decrease due to the INVERSE law of Sound and atmospheric conditions. If the two people were only centimeters apart and talking on axis's to the other's ear, the high and low frequency content of the signal would sound enhanced and emotionally intimate if speaking softly. I will be discussing in detail regarding optimizing the dimensional effect of using direct path sound later in the commentary.

Precedence Effect

In a typical listening situation, the listener receives the direct sound emitted by the sound source as well as delayed and attenuated versions of the direct sound resulting from the reflection of the sound in the environment. The reflected sounds reaching the listener may emanate from any direction in the environment, potentially creating a false impression of a sound source at the location of reflection. However, this is certainly not the case as the auditory system can clearly localize a sound source in the presence of multiple reflections (reverberation). The ability of the auditory system to "combine" both the direct as well as reflected sounds such that they are heard as a single "entity" and localized in the direction corresponding to the direct sound has been termed the precedence effect also known as the Haas effect and the law of first waveforms. The precedence effect allows us to localize a sound source in the presence of reverberation, even when the energy of the reverberant or reflected sound (Delay) is greater than that of the direct sound (See Above).
Various experiments to investigate the precedence effect include a listener and two loudspeakers, placed in a triangular setting, in an anechoic environment. One loudspeaker is used to provide the direct sound while the other provides a delayed and appropriately attenuated version of the direct sound in order to simulate a reflection. Such studies indicate the following:

1) When the reflection and direct sound are generated simultaneously, A single sound source (virtual source) is perceived at a location half way between the two loudspeakers. (Phantom Centre-Mono)

2) As the time delay between the direct sound and the delayed sound is increased from 0-1ms the location of the perceived sound source moves towards the "direct sound loudspeaker" (this is known as summing localization).

3) When the delay is between 1 and 20msec the sound source is correctly localized (e.g. coming from the direct sound loudspeaker) without being affected by the reflected sound.

4) When the delay exceeds approximately 20msec, the direct sound is correctly localized, however, the delayed sound is also localized as a distinct sound source at the position of the “reflection loudspeaker”

5) If the delayed sound source is delayed approximately from 1-15msec from the original and is slightly louder than the original sound source, the listener will perceive the sound source location as coming from the original sound source even though it is lower in amplitude.

The experiments show how we are capable of correctly localizing a sound source in the presence of reverberation provided the reflections arrive within a short period after receiving the direct sound. The possibilities of using the precedence effect in widening the image of the dedicated centre speaker will be explored further in the commentary. (See Fig:

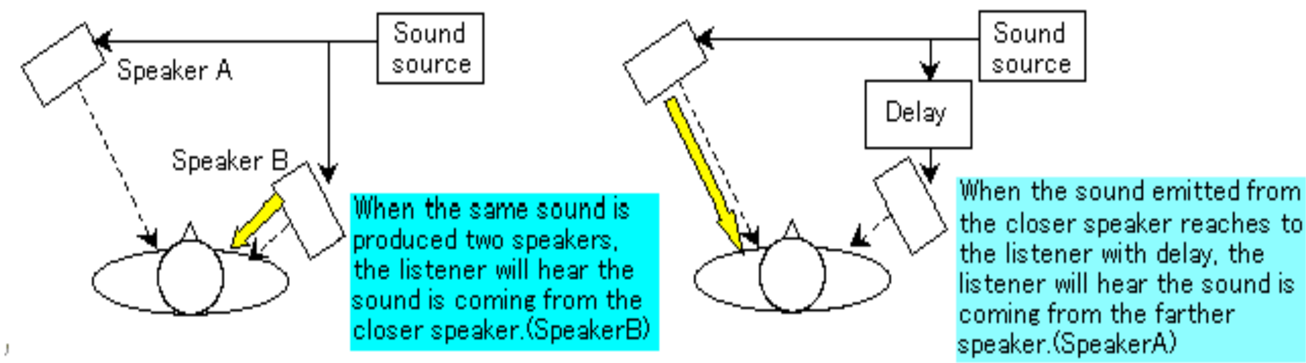


Fig: 2 The Precedence (Hass) Effect

First and Early Reflections

Reflections of order one, resulting from the room boundaries (e.g. walls, floor and ceiling), are known as early reflections and typically arrive approximately within 20–80msec of the direct sound and will differ in amplitude and frequency content. The greater the time difference between the early reflections and the direct sound will be mirrored in amplitude and frequency content differences with the amplitude and high frequencies dissipating exponentially over time. With reflections arriving at 60msec and 80msec, the sound created will dictate that the reflective surfaces are at a greater distance than reflections arriving at 15msec and 30msec. Reflections arriving less than 10ms will produce a flanging or phase effect if the walls are parallel and perpendicular to each other. (See Precedence Effect) This effect can be easily produced by one's clapping of hands and listening for a flutter echo flange with itself-which is caused by multiple reflections arriving less than 10ms from each other. Once the first and early reflections pass the 80msec mark, they begin to sound discrete from the original signal and do not contribute much in influencing a sense of distance in the overall sound. Later we will look at how first and early reflections can play a role in creating dimension in surround sound mixing.

NB: The transient nature of the original signal will influence the 20msec--80msec range of early reflection properties. A transient snare drum signal may begin to sound discrete at the 50-60msec point, where a smooth signal like a Cello will not begin to sound discrete from the original until 100msec.The amount of high frequency content at the front of the signal's waveform also is a influencing factor.

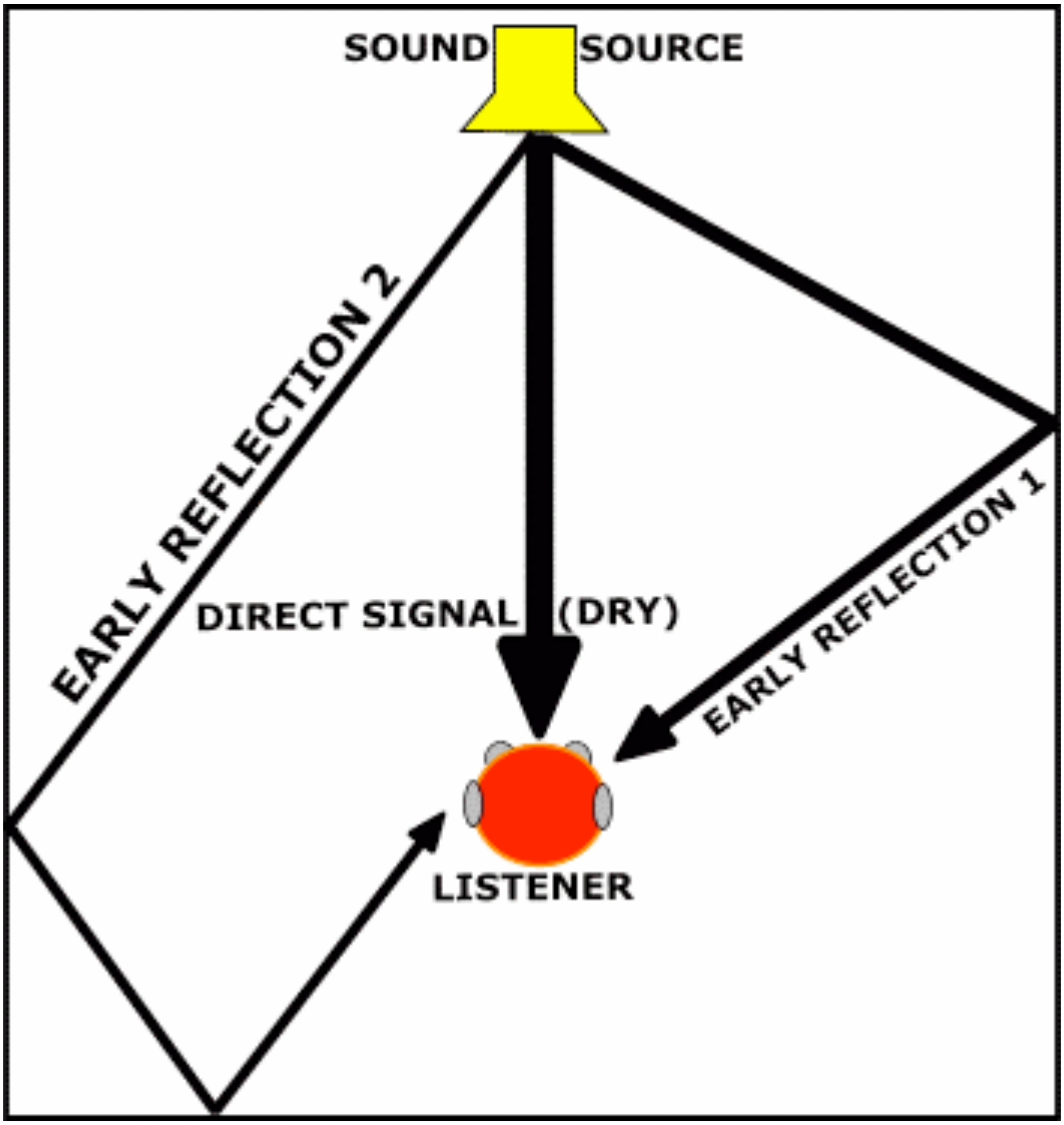


Fig:3 Direct Sound and Early Reflections

Reverberation

When sound is produced in an enclosed space, multiple reflections increase and blend together, creating reverberation/reverb. This is most noticeable when the sound stops, but the reflections continue, decreasing in amplitude, until they can no longer be heard. The time it takes for the sound pressure level of the reverberation to decay 60 decibels is known as the reverberation time, or RT (60). (Wikepedia)

As shown in Figure 3; In a typical listening environment, sound waves emitted by the source reach the listener both directly, via the straight line path between the source and receiver and indirectly as reflections (e.g. echoes) from any walls, floor, ceiling or any other obstacles and obstructions. This collection of reflected waves, which may consist of several thousands, reflecting from the various surfaces within a space, is known as reverberation.

The collection of reflected sound reaching the listener varies as a function of the geometry of the room relative to the listener, as well as the material of the room (absorption coefficients) and the frequency components of the source spectrum. Reverberation can also be used as a cue to source distance estimation, and can also provide information with regards to the physical “make-up” of a room (e.g. size, types of materials on the walls, floor, ceiling).

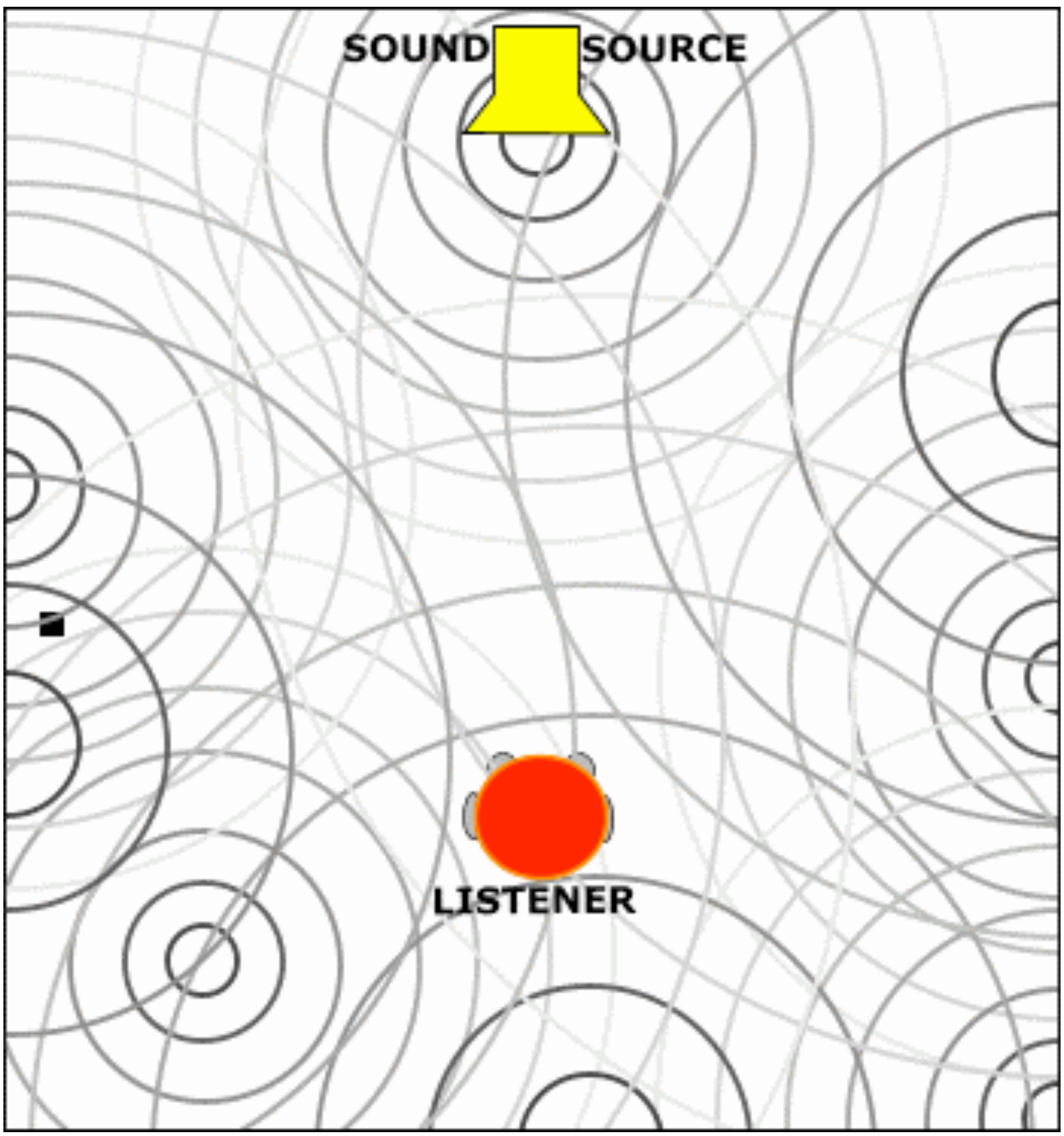


Figure 4: Reverberation

The number of times a wave is reflected before reaching the listener is known as its order. The direct sound has an order (one) of sound arriving once to the listening position. In a typical scenario, the number of reflected waves may reach several thousands. A reflected wave is denoted by its order of multiple reflections.
In many situations, a higher reflection order indicates a reduction in the intensity level due to the absorption by the reflecting surfaces and the inverse square law characteristics of propagating waves.

In addition to the direct sound, reverberation can be broken down into two categories: early and late reflections. Reflections of order one (Direct Path), resulting from the room boundaries (e.g. walls, floor and ceiling), are known as early reflections and typically arrive within 80ms of the direct sound and will differ in amplitude and frequency content. The greater the time difference between the early reflections and the direct sound will be mirrored in amplitude differences with the amplitude dissipating exponentially over time.

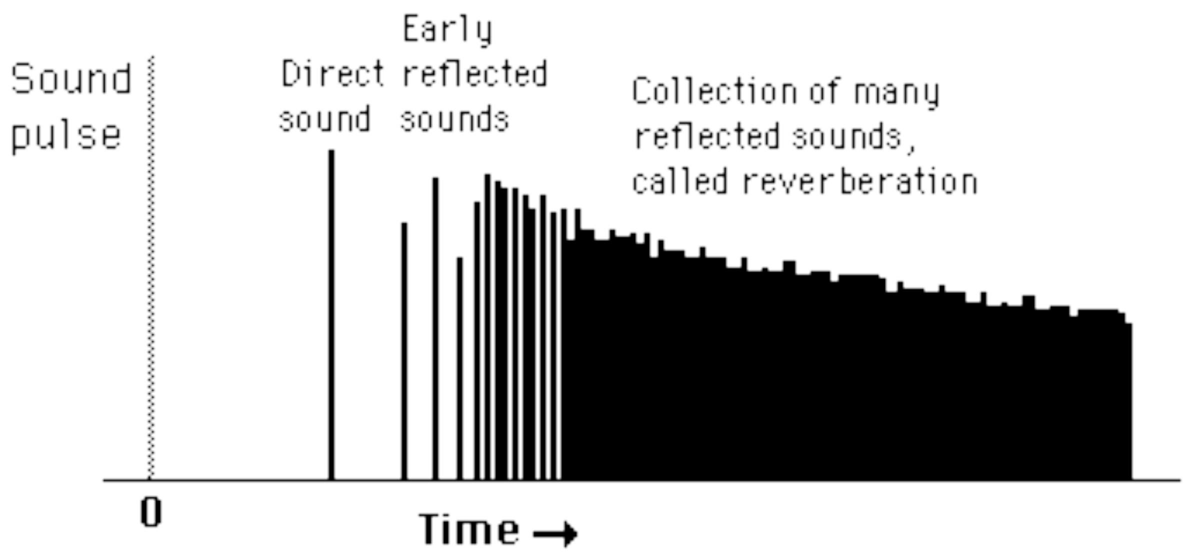


Fig 5: The 3 Components; Direct Sound, Early Reflections and Reverberation

The Reflections arriving after 80ms and with reflection orders greater than one are known as late reflections or better known as reverb or discrete delays.
As the direct path sound decays, the initial sound of the late reflections and reverb will some times be louder than the decay of the direct sound, thus sounding enmeshed or even detached from the original. Late reflections, arising from “reflected reflections” from one surface to another, are assumed to arrive equally from all directions and even amplitude to both ears (e.g. diffuse) and can be described statistically as exponential decaying sound (RT-60)(See Fig 5)

Reverberation time T60 can be defined as the time required for the sound pressure level (SPL) to be attenuated by 60dB (e.g. by a factor of one million), independent of the intensity of the sound after a steady state sound is turned off and can be approximated by reverberation time, as given, is rather arbitrary and depends on the characteristics of the enclosure, including the material of the walls, floor and ceiling, number and type of objects in the room etc. Depending on the level of the background noise, it may be the case that reflections arriving after T60 are still considerably audible. However, the choice of 60dB was made by considering a good "music making area," such as a concert hall. In such a situation, the loudest level reached for most orchestral music is typically 100dB (SPL), while the level of background noise is around 40dB. As a result, a reverberation time of 60dB can be seen as the time required for the loudest sounds of an orchestra to be reduced to the level of the background noise.

Reverberation time is highly affected by the reflective surfaces encountered by the propagating waves. When a surface is highly reflective, very little energy is absorbed by the surface (e.g. the reflected wave contains most of its energy) leading to an increase in the reverberation time. In contrast, highly absorbing materials will absorb much of the energy of a wave striking it, greatly reducing the energy in the reflected portion thereby reducing the reverberation time.

Late reflections can be considered diffuse, however, as the distance between the source and listener increases, the intensity (loudness) of the direct sound will decrease until the level of the direct sound equals the level of the reverberation reverb.

It is important to note, that after a signal stops emitting audio, the reverb continues for an RT-60 that is indicative of the environment and can be last up to times over ten seconds. If a hall was filled with highly reflective surfaces only and had no openings for the sound to escape you could theoretically create an effect where the signal may last forever, for it is energy we are dealing with. However that type of situation would be impossible to construct. But what we must look at is what happens to the reverb signal while it is decaying. If you analyzed the frequency response of a reverb signal at the 1-second mark and then at the 3-second mark, you would notice that the mid-high frequency content of the signal would decrease as the reverb amplitude decreases. The amount of loss of mid-high frequency content would be determined by the absorption coefficients of the reflective surfaces.

Reverberation can add a pleasing aspect to voice and music, making it attractive to a demographic that prefers harmonic and melodic content over rhythmic content.

Many of today's home theater manufacturers have also taken advantage of the benefits reverberation has to offer. Many radios, sound-systems, and home theater systems include DSP technology offering various reverberation settings.

Greater details regarding the characteristics of reverberation are provided in the following sections

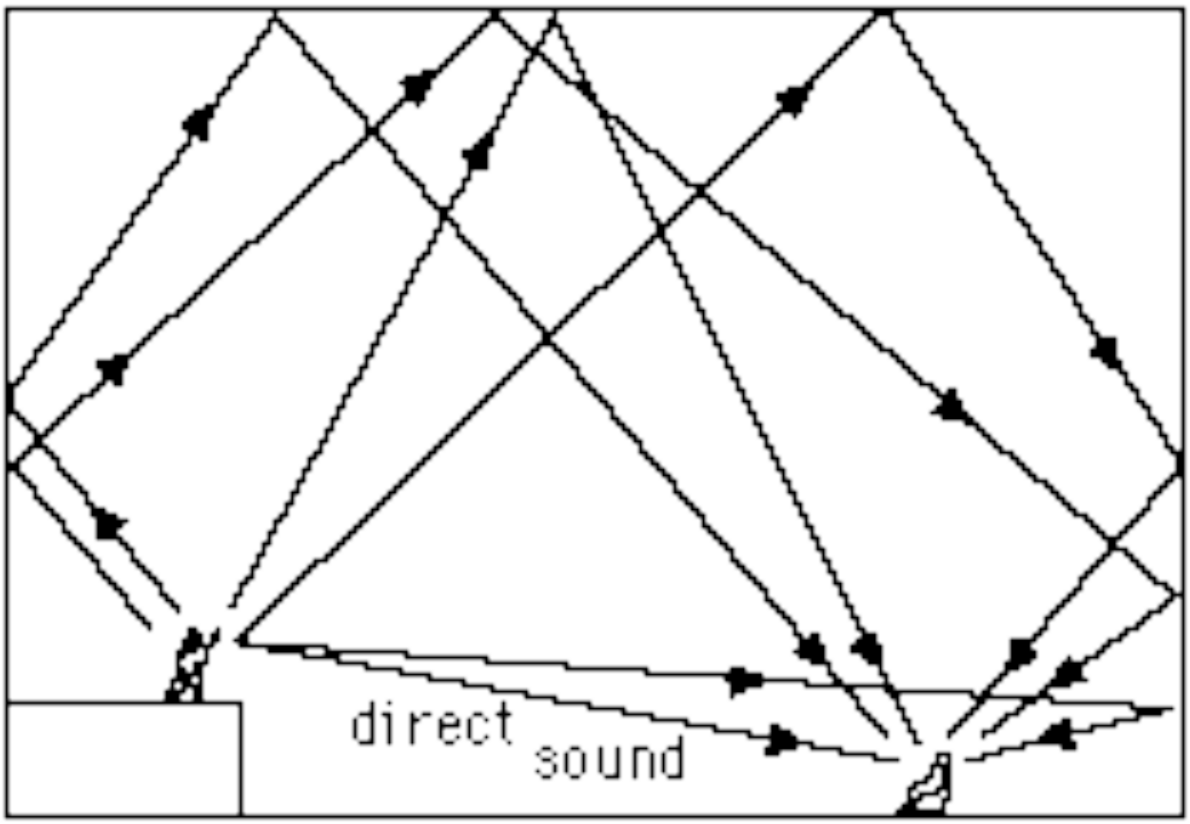


Fig: 6 Direct Sound, Early Reflections and Reverb-Concert Hall

Auditory distance cues

The following auditory distance cues may potentially play a role in the perception of the distance to a sound source when both the observer and the sound source are stationary:

1. Intensity (sound level) of the sound waves emitted by the source.

2. Reverberation (direct path-to-reverberant energy).

3. Frequency spectrum of the sound waves emitted by the sound source. Auditory distance will be emphasized when the difference in arrival time to the listener's ear and frequency response of the direct path sound and the reflected reverberant sound.

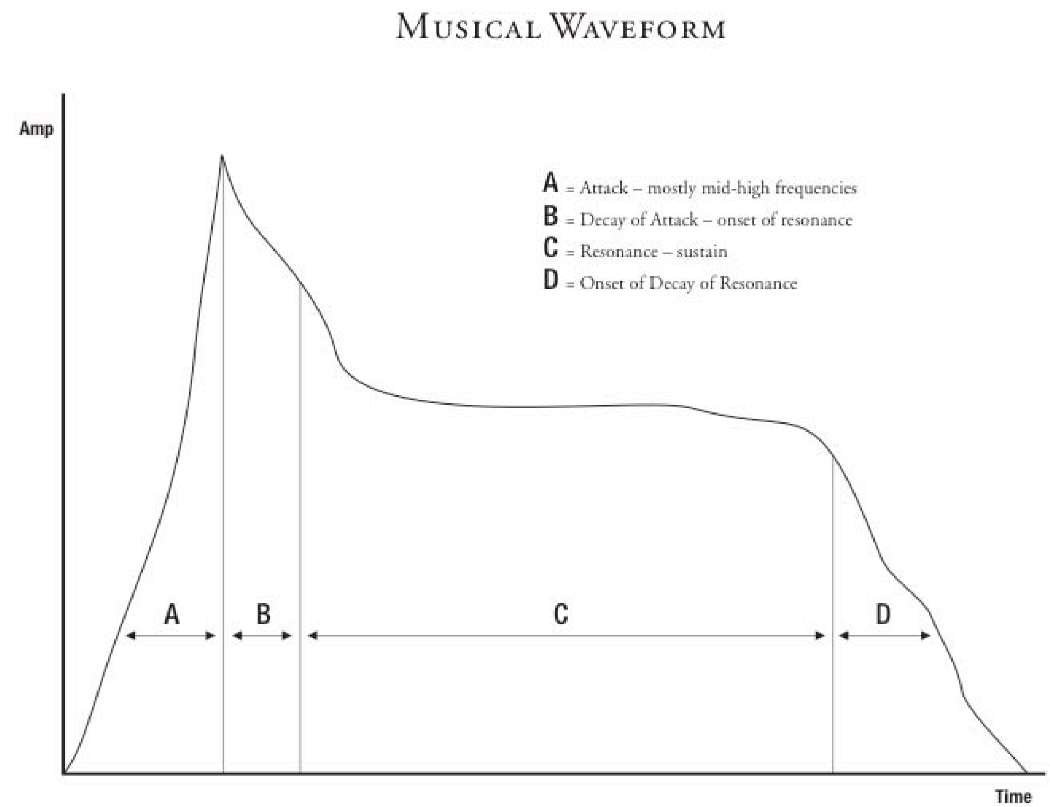
4. Binaural differences (e.g. ITD and ILD).

5. Type of stimulus used (e.g. familiarity with the sound source).

Source intensity (sound level) and reverberation (direct-to-reflection/reverberant energy) are believed to be the most effective factors in determining distance between the originating sound source and the listener, however, any number of these examples may be present and certain examples may dominate depending on the listening environment. As a result, auditory distance perception may be influenced by such factors as the user's familiarity with the room's reflective properties as well as the stimulus and the distance estimation process actually employed by a listener. In addition, changes in these cues may not necessarily be due to a change in distance between the listener and the source, but rather, may result from changes in the spectrum emitted by the source (e.g. the source power is reduced), or changes to the source spectrum due to changes in the environment, thereby further complicating matters, leading to poor judgments in source distance estimation. For example, as source distance is increased, the intensity of the sound received by the listener decreases. However, sound source intensity of the sound waves received by the listener may also decrease without an increase in source distance, but rather with a decline in source intensity.
In such an ambiguous situation, the user may not necessarily be able to discriminate between the two scenarios. Fortunately, as described below, the presence of other distance examples may assist the listener in making the correct judgment.

It appears that auditory distance studies should be conducted in normal, reverberant environments. Source distance cues can be divided into two categories, exocentric, and egocentric. Exocentric or relative cues provide information with respect to the relative distance between two sounds whereas egocentric provide information about the actual absolute distance between the listener and the sound source. Consider a sound source and a listener in a room where the listener cannot see and does not have any prior information regarding source position or distance. Now further imagine the source distance is doubled. Using the decrease in sound intensity between the sound source at the initial position and the sound source at the new position to determine that the distance has increased, is an example of an exocentric cue.

On the other hand, when the listener uses the ratio of direct-to-reverberant levels to determine the source is five feet away from him or her is an example of an egocentric cue.



The Waveform

Important to optimizing one's creative input into Surround sound production is the knowledge of the audio waveform; amplitude-dynamics, time duration and frequency content. There are four sections of the waveform to analyze in detail in how one can alter the original waveform to a desired waveform for their creative purposes.

(A)-The Attack

In most audio waveforms, the attack *(A-section)* is mostly made up of mid-high frequency content with little mid-low frequencies that are associated with music fundamentals. When analyzing the waveform of a piano chord, the first sound one would hear is the attack of the hammers hitting the strings producing high overtones. This attack would sound very percussive and almost noise-like if it is heard on its own.

Once the strings have been struck, they start to vibrate producing some sustaining musical elements (*B-section).* When the strings are vibrating they start to excite the soundboard producing musical notes (*C-section).* After the player stops playing, the piano will still produce sound momentarily (depending whether the sustain pedal is being utilized) until all sound decays. With a drum, it will be when the stick first hits the head. With dialogue, words that begin with hard consonants, there is no tonal content in this part of the waveform and what is present is a signal containing sonic elements that are similar to noise properties. In dialogue with words like ‘Time’ the ‘T’ part of the word contains mostly noise. The “ime” part of the word contains tone-the vowel sound with pitch. It is safe to conclude that when editing dialogue you can literally take any word that begins with ‘T’ and by its sonic character, use it in other places in the dialogue that contain ‘T’s at the beginning of a word. This is not true for vowel sounds like the ‘ime’, for it will contain a certain inflection associated with pitch.

With music, the *A-section* defines the rhythm-the attack. If the instrument is a piano playing eight notes with sustain, a situation might arise where you wanted to focus on having the piano provide more rhythm than harmonic content to the production.

In the waveform of a piano chord, the attack (*A-section)* contains more mid-high frequency content than the sustain (*C-section)*. As the piano chord sustains and decays, so do the mid-high frequency and amplitude components in relation to the attack (*A & B-sections)* of the chord . When the piano is played with dynamics, the attack (*A-section)* of the signal varies in mid-high frequency content in relation to amplitude. The harder the attack-the brighter the sound. The difference in frequency content and amplitude while the chord sustains (*C-section)* does not change as dramatically when there are minimal dynamics in the attack (*A-section)*. The differences in attack amplitudes only subtly influences the frequency and amplitude of the sustain. (*C-section).*

If you desired to create a rhythmic element instead of a harmonic element in a song with the piano playing eight notes with sustain you could do so with the following technique;

Compress the piano with a med-slow attack time and med-slow release time to elevate the level of the attack (*A-section)* in relation to the sustain (*C-section).* Increase the mid-high frequencies so the attack (*A-section)* is defined and articulate.

You need to compress the piano before you increase the mid-high frequencies, because you do not want the sustain section of the piano to be as bright in relation to the attack. If the mid–high frequency elements of the sustain (*C-section)* section are always reflecting the increase in the attack equalization (*A-section)*, than the difference in mid-high frequencies will not be enough to distinguish the rhythmic element from the sustain element of the piano chord, which could create problems with the melodic elements such as a lead vocal that will often need to sound present in a mix. All you are doing here is equalizing the entire compressed waveform of the piano. The idea is to enhance the attack section of the waveform only to assist in highlighting the rhythmic elements of the piano performance.

To emphasize the sustain element (*C-section)* of a piano chord, the process would be different in the dynamic processing of the signal.

Often in productions other musical instruments will provide the musical rhythmic function instead of the piano such as a chicken picking guitar part, where the piano would now be used in supplying the main harmonic foundation in the production. If this is the case, one needs to enhance the sustain (*C-section)* of the piano.
Compress/limit the piano with very fast attack and release times. This will lower the attack (*A-section)* in relation to the sustain (*C-section).* Once this has been achieved one can then equalize the sustain (*C-section)* in the frequency area, 200hz-1khz for the harmonic component, and 2khz-5khz for mid-range presence. If one where to remove all frequency information between 200hz-1khz from a production, they would be left with s production containing only bottom and mid-top end, with no organic impression of harmonic and melodic music (music range).
After compression, one could enhance the music range to provide more harmonic structure than rhythm.
To conclude, most instruments have the ability to provide a combination or singular design of rhythm, harmonic and melodic ideas. A lot of this can be achieved by synchronizing the right instruments with the right parts, or through the manipulation of it's innate waveform .

> NB: With the sustain part of the piano (*C-section)*, the difference between high frequencies (10khz-15khz) and mid-range frequencies (2khz-5khz) is substantial. If one wanted to enhance the presence of the piano, they would have to increase the amplitude of the high frequencies a great deal more than the mid-range frequencies to achieve the same presence effect. If one were to go this route, the lead vocal will sound dull in comparison to the piano, where to

compensate, one would have to increase the high frequencies of the lead vocal for it to stand out. By doing this, the musical component (200-1.5khz) of a production will sound very detached from the sonic components (3khz-15khz).
This is a common problem in today's music production, where the sonic elements of a song are very detached from the musical elements, through excessive equalization and dynamic compression. I equate it to listening 10metres behind a 747 Jumbo jet-*loud noise with no musical content!*

(B) The decay of the attack-Onset of Sustain

This part of the signal is a mix of the decay of the attack and the onset of resonance and pitch (*B-section)*. As the attack part decays, the first sign of pitch begins. The change from A to B and occurs rather quickly and is not noticeable to the average human ear. This is the suggested point in a drum waveform for sample enhancement.
Take your favorite drum sample and remove the attack (*A-section)* of the waveform and trigger the sample with a gated–limited key input from the original drum. This will allow you to retain the attack of the original drum and enhance the overall sound with the sample. The sample can create an idea of greater size to the overall sound through duration instead of amplitude, and it most cases the amplitude of the sample is never as loud as the original's attack.

(C) Resonance/Sustain

This part of the signal is the sustain part that contains the resonance and pitch of the signal (Music). This is where vibrato and tremolo would occur. The sustain (*C-section)* part of the signal is where compression is used to control an overall volume adjustment to minimize dynamics so a signal can be heard clearly.

(D) The Decay

This part of the signal is the decay part that occurs when the audio stops projecting. Most of the audio content of the decay part is the reflective sound of the environment. It concert halls this can be as long as 4 seconds and shorter than a quarter of a second for an interior room environment.

Recording

A conventional approach to surround recording is to use a method to record with the goal of emulating the natural acoustic environment. An alternate approach is to abandon the conventional rules to enhance a creative approach. A third method would be to use components from both methods.

The traditional and most common approach to 5.1 for music is with orchestral recording. Sound design, ambience, and dialogue are mostly mono or stereo elements processed for creating a surround image.

Orchestral recording is often the only component truly recorded and mixed for surround. One standard process is to use a Decca Tree where the recording engineer will place 3 microphones (same model) in a configuration directly over the front of the orchestra above the conductor to be panned over the front channels-Left-Centre-Right. This pick-up captures what the orchestral balance as heard by the conductor perspective.
On most recordings the pick up pattern of the microphones is set to omni, where the height dictates the latitude of the stereo image and the mix of direct path sound to the early reflections/reverb sound.
However it is limited for it only satisfies a stereo perspective and any surround enhancement would have to be manufactured by the engineer. (See Fig:7, Fig:8)

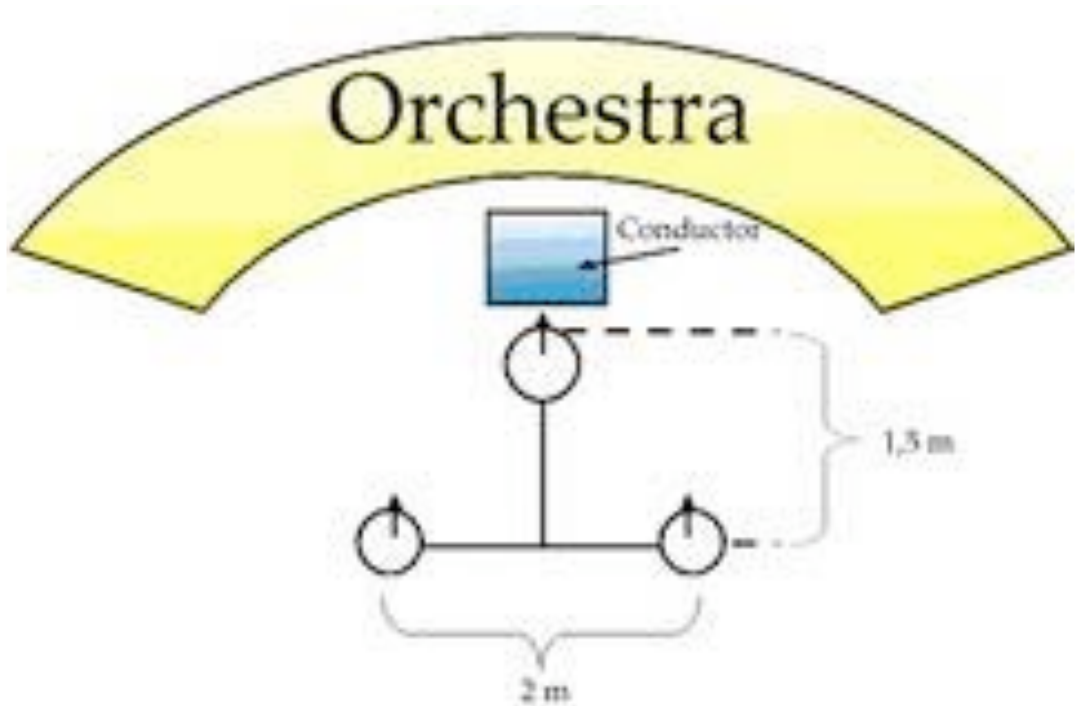


Fig:7 Decca Tree 3 Microphone Configuration

Fig:8 Decca Tree

Another approach for recording surround sound is to use the Decca Tree with flank microphones on either side of the orchestra, where the recording engineer will place 3 microphones in a configuration directly over the front of the orchestra above the conductor to be panned over the front channels-Left-Centre-Right. The left and right flanks microphones are panned hard left and hard right and allow the engineer to increase the perspective of the stereo width. This pick up pattern allows the engineer to position the Decca Tree lower to the orchestra for a tighter pick up whereby the flank microphones pick up more of the room sound along with a wider stereo capture. On most recordings the pick up pattern of the microphones is set to omni and like the 3 microphone Decca Tree this pick up only satisfies a stereo perspective and any surround enhancement would have to be manufactured by the engineer. (See Fig:9)

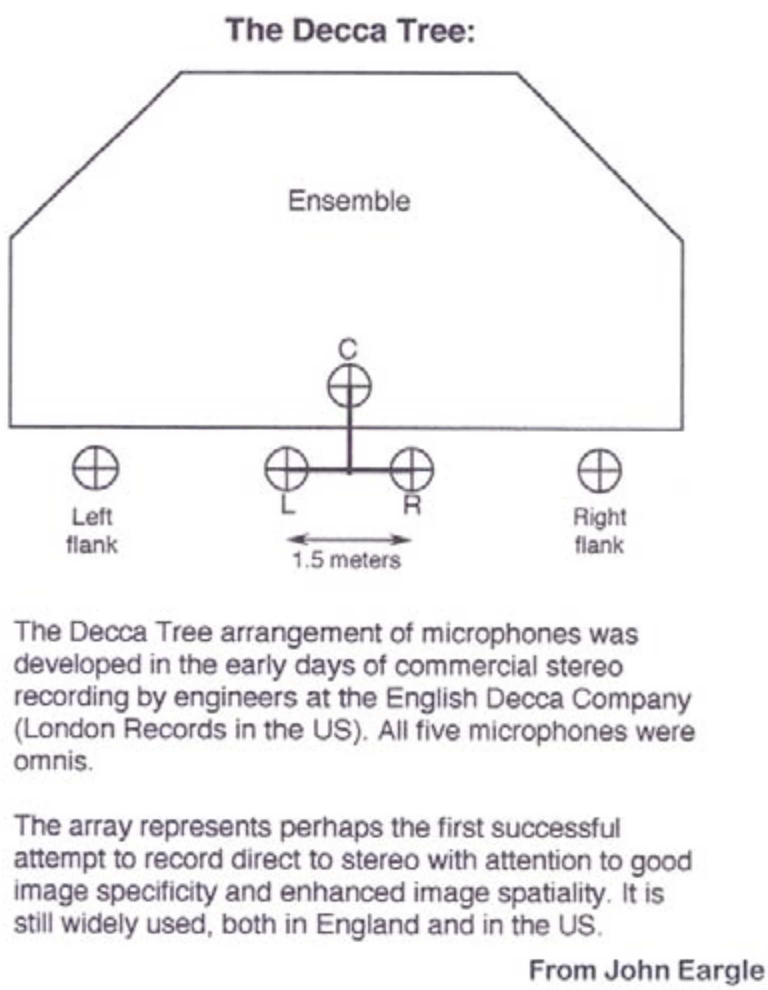


Fig:9 Decca Tree with left and right flank microphones

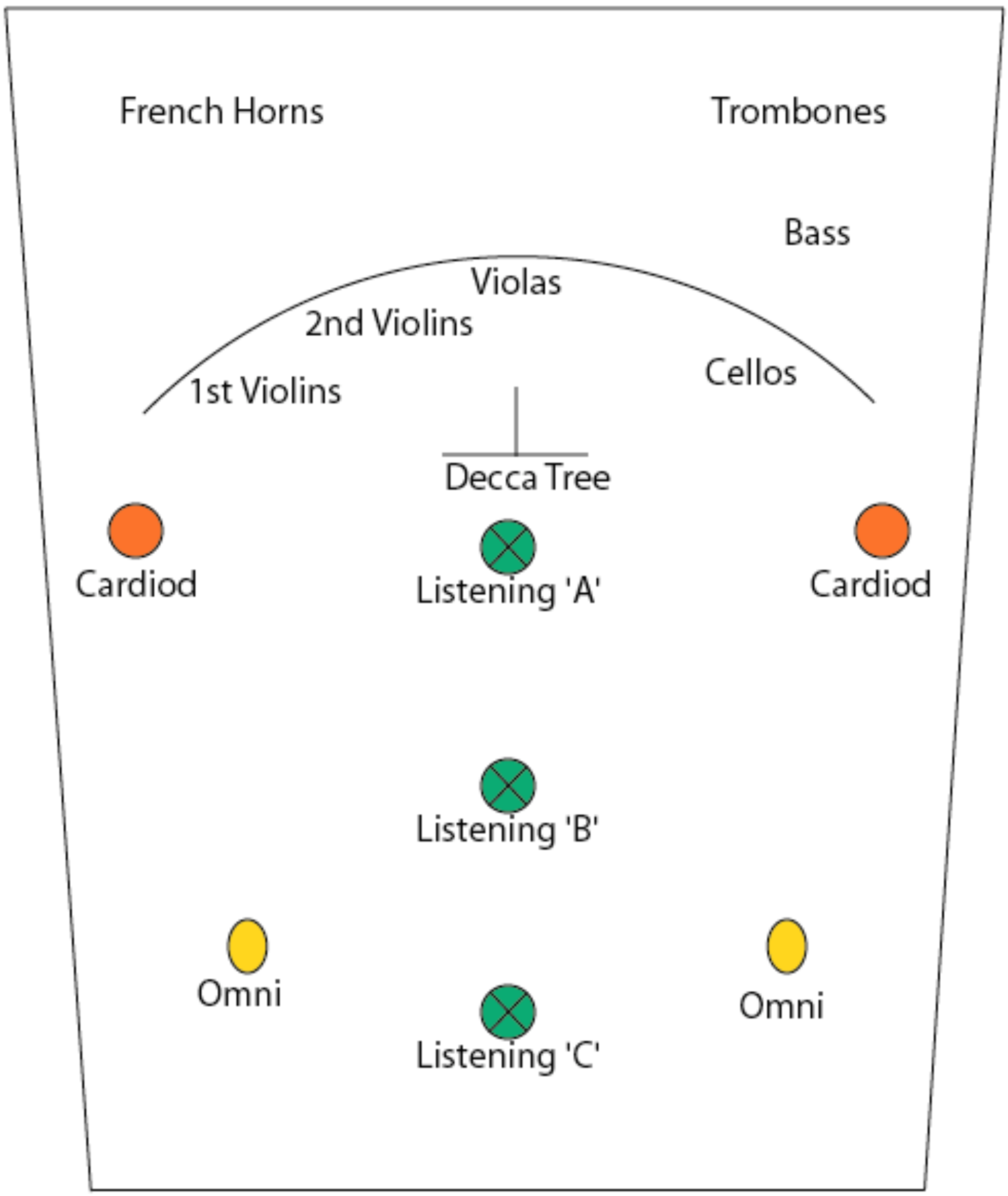


Fig:10 Decca Tree, Flank Microphones and Ambient microphones

Fig:11 Decca Tree with Left and Right Flank Microphones

With music playing a more important role in film entertainment today, many executives will allocate more money in the music budget to allow for a true surround pick up. This type of recording requires a large space (sound stage) with excellent ambient reverb qualities. However sound stages require a lot of real estate and are often only found in the larger film centers of the world.

The standard surround record used around the world today integrates a combination of pick ups, so the engineer can have control over the balance and overall sound of the orchestra in the mixing stage.
An engineer will use position 20-30 spot or close microphones on certain instrumental sections in case the composer wants to feature a certain instrument in the mix.
In addition is the Decca Tree, 2 flank microphones and 2 ambient microphones. The ambient microphones are always omni as are the flank microphones and the Decca Tree.

The panning configuration is:

1) Close microphones positioned across the front channels, panned to where the section would appear in a left to right image when one is looking at the orchestra from the front.
2) The Decca Tree and the flank microphones are placed across the front channels with the centre microphone of the Decca Tree panned to the discrete centre channel.
3) The ambient (rear) left and right microphones are panned to their rear channels respectively.

The spot microphones are only used if the composer needs to bring up the level of a certain section in the score, in case the actual players did not produce enough volume or the rest of the orchestra is to loud and washing that instrument out.
The Decca Tree focuses on the conductor’s perspective. The flank microphones add more width to the image of the orchestra with additional ambient sound from the sound stage. The engineer will decide if the flank microphones should be in cardiod or omni patterns depending on how much of the ambient sound he wants in those microphones. Also a factor that might influence the engineer is the fact that a microphone in an omni pattern has a more even frequency response pick up than the same microphone in cardiod. Something to consider if quality is paramount.

The ambient microphones are used to pick up the reverb characteristics of the recording space. They are positioned equal distance from the back and side-walls in order to pick up maximum diffusion without early reflections influencing the over all pick up. The microphones are usually small diaphragm condensers and are the same model so each channel sounds very similar in characteristic.

If one is to apply the different microphone pick ups to the diagram of the waveform, one would conclude;

1) The close microphones would feature the A section of the waveform.
2) The Decca Tree and the Flank microphones would feature a more equal combination of A-B -C sections.
3) The ambient microphones would feature the C-D section of the waveform.

Most production teams on a surround record session aim for an overall combination of the various microphone pick up's in order to arrive at a mutually satisfactory sound that indicates what balance of sections of the waveform they prefer in the final sound. Once this is achieved it is left as a preset for the entire music score.

Release the Hounds !!!!

<u>I believe this type approach to surround recording and mixing is limited.</u>

Although most engineers prefer an overall pick up that emulates the sound stage they rarely add in the section microphones to the final mix and reluctantly will if the composer needs a certain section to heard specifically. The final product then sounds like the listener is situated at a fixed distance from the orchestra. This works well in theory, but at times it does not translate very well to the optimum listening position in the theatre (sitting approx 1/2 to 2/3 from the screen). One problem with this approach is that there can be a lot simultaneous audio going on in the film from effects, dialogue and sound design, that you barely notice the rear channels due to the masking effect and lack of rhythmic articulation. Even when there is only music in the mix, the engineer will maintain the Decca setup across the front channels, with the ambient microphones in the rear channels, occasionally turning up the volume of the rear channels to create more of a true realistic hall effect. What might occur when this happens?

With a fast tempo, the music sounds harmonically undefined and inarticulate for the ambience starts to mask the beginning of the next music envelope. Relating to the above diagram, the 'C' part of the waveform elongates, gets louder and will overpower the 'A' part of the waveform of the next incoming signal. Considering that the rhythmic characteristics of music come from the 'A' part of the waveform, you can certainly see how the rhythm can sound obscure for the buildup of the resonance, 'C', will now mask the rhythmic interpretation of the composition. If the envelope has a slow attack time the problem will be exacerbated.

In some theatres I've noticed harmonic dissonance due to the fact that a lot of the ambience of the hall microphones, added in with the theatre's RT-60, (contributes to a longer reverb time for mid-low frequencies in the theatre, creating an effect somewhat like a piano player playing at a fast tempo with the sustain peddle down all the time. In a good recording hall you will get at an approximate reverb time of 1-2 seconds that will be captured by the rear microphones. Add that music pick-up in with a large reflective theatre you may get a T-60 time of close to 2-2.5 seconds in the lower frequency range. This might sound great for Adagios but once the tempo is picked up the music will soon sound harmonically confusing once it arrives to the listeners ears. Also when the music is constantly at a low level to allow room for dialogue, the first thing to suffer is the A section-articulation of the music.

As earlier stated the A-section contains most of the music's rhythm (mid-high frequencies). Not only that, because the close microphones are situated closer to the instruments than the Decca Tree and ambient microphones, the signal present will be the earliest of the 3 microphone pick ups; if we looked at all 3 waveforms in a DAW like Pro-Tools. Also the level of the A-section of the close microphones will be lower in volume than the C-section of the ambient microphones.
But conventional production crews rarely use the close microphones in the final mix or alter the balance between the Decca Tree, flank and ambient microphones to correct harmonic clutter or enhance rhythmic clarity.
No wonder the sound gets washed out at fast tempo.

When I mix music for films, I will do split mixes where I have the close microphones mix on one stem, the Decca Tree and flank microphones on another stem and the ambient microphones on another. This allows me to balance between the articulation (A-section) and the sustain (C-section) parts of the music in the final mixing stage. If the tempo is slow, I can add in more of the flank and ambient microphones and increase and elongate the surround reverb time to fill out the composition with more harmonic sound duration (C-section). If the tempo is brisk I will get a balance of the close microphones (A-section) and mix them into the final mix at a level where the rhythmic articulation is clearly heard on the sections required. I can also decrease the amplitude and time of the surround reverb if I am using it. Remember, all I need is a small amount of level from the spot microphones to emphasize the A section of the waveform. If the music's focus is more on its rhythm than harmonic structure and is secondary to the dialogue, mixing in the close microphones will allow for clarity in the music when the music is mixed in at a lower level compared to the dialogue. Does the overall sound change? Yes, but not enough to notice. I am only adding in the mid-high frequencies of the A-section

of the close microphones so the harmonic elements of the overall mix remain almost the same. What one might notice when listening in a theatre, is the aural suggestion that the seating position has moved slightly closer to the actual orchestra. A small price to pay for needed articulation in the overall sound.
If the tempo slows down, all one has to do is reverse the process by adding in more of the ambient microphones and extending the surround reverb time.

In rock and pop music the tendency to create a natural environment is desired but also creative sound sourcing. Placing instrumentation in various locations can prove top be exciting for the listener. One could only imagine how *"Dark Side of the Moon"* would sound in surround. Placing instrumentation in various sound locations can also prove to be distracting if there is a main focus going on like a lead vocal or a solo. Listening to a lead vocal while the high hat panning in the rear speakers might sound cool but will obviously pull focus from the lead vocal. Having the lead vocal coming from the front and the solo from the left will prove to be interesting and should not cause the listener to lose focus as long as one of them is not being panned. I always look at it this way, maintain focus with as little distraction as possible. Remember the human ear works similar to vision; you can focus on something straight in front of you, but as soon as something enters your peripheral vision the eye will change focus to what is moving.
By all means place sound sources where you like, but make sure your production maintains focus.
Recording every instrument in surround would be very difficult and limiting; everything would sound the same dimensionally and there would be no sense of depth to your production. It would be all right for a live recording or an orchestra; recording, but for pop and rock it will most likely sound boring.
What one good to for example is to record the instrument with a stereo ambience; a close microphone on the source and a stereo microphone for the sound of the room. You could pan the close microphone left middle and the stereo room front left and rear left. This would give you the image the instrument is coming from the middle of the left side with a room sound coming from the left front and left rear. You could do this with an instrument for the right side and then you would discover that instead of instruments coming from mono sound sources, there would be ambience with them giving the idea that there is depth to the sound. Add in stereo reverb and stereo delays in addition to the stereo ambience and you could really create the idea that an instrument is coming from a room beside you or even behind you.
Remember your production needs to have the space for you to hear the ambience. If the instrumentation is dense in your production you might one to create dimension using mono imaging only.

If then instrument is the lead vocal and it is panned centre, you might want to create the effect where they are singing in a concert hall with you sitting a couple of rows back. A surround sound reverb is very effective for this when configured correctly. Reverb time, equalization, pre-delay, early and late reflections (manufactured by DDl's) play pivotal rolls when getting a surround reverb to sound natural and convincing. If you do not have access to a surround reverb, you can create one by using 2 different digitals reverbs panned across the front and rear channels. The digital reverbs have to be different plug ins, for if you assign a signal to 2 digital reverbs with the same algorithm, the reverb signal will collapse into stereo between the front and back channels. More about this later.

Breakdown of an Audio Signal in a Closed Space

When it comes to creating the impression of a believable reverb environment, what are the factors that contribute to achieving this?

An audio signal takes three different paths while listening in an enclosed environment.

1) The direct path signal from the originating source to the listening position
2) The first and early reflections coming off the walls, ceiling and floor from the source to the listening position.
3) The many diffused reflections know as 'reverb' arriving to the listening position. The individual level and frequency response of these signals determine the size and the quality of the listening environment.

The unobstructed direct signal is always the loudest and most defined in its frequency response. The time it takes for the signal to travel from the source to the ear is determined by the speed of sound (approx 1 meter/sec). If the audio source moves slightly to the left, the ear will distinguish this movement for the audio signal will arrive to the left ear slightly sooner than the right ear (ITD). As distance is added the sound loses amplitude and its frequency range decreases because of atmospheric conditions so the ear acknowledges the sound source is further back.

The first indication of dimension is when a reflected signal arrives later than 15msec but before 100msec from the original at a lower level. If the reflection arrives before 15msec it won't create dimension and instead imaging problems. In most listening environments there will be 2 delays called the first reflection, coming from the left and right walls. In most circumstances the delay times will be slightly different from each other but distinct from the direct signal. The delay's frequency range is always smaller than the direct path signal and the amount of reduction is based on the reverb coefficient properties of the reflective surfaces.

For example: If you are sitting 5 meters from the sound source exactly between two walls 25 meters apart, the direct sound will arrive in 5ms and the reflections from the walls will arrive in 25msec, a difference of 20msec. If you move slightly off centre the left reflection will not equal the right reflection; the earliest reflection will come from the closer wall and will be louder and ever so slightly brighter but that noticeable to the average ear. The reflections will be lower in level and contain less high frequency, for the walls will be absorbing some of the sound.

In a mix if the left and right delays are exactly 36msec and are identical in sound, you will determine that the direct sound and the delays will be coming from the same place and that it will sound mono, which is unrealistic in a natural listening environment for it is impossible to have the left and right delays arriving at the same exact time with the same amplitude and frequency response. So to create dimension in an environment like this, one needs to take liberties with the delay settings.

First the goal is to create distance with a localized direct sound image. To do this, place the original signal in the centre position and incorporate the first of the 2 delays at least 15msec from the original signal to prevent any phasing or flanging effect. Have the original signal panned centre and add in a delay panned centre (use low-pass filter) at a time setting and level to create dimension. Remember to roll off some high end from the delay, for reflections are as never as bright as the original signal. If you decide 40msec at -6db from the original signal is what you prefer, then work with that sense of dimension. However just one mono delay will not create stereo dimension! For believable dimension in stereo you need 2 dealys.There also needs to be at least 15msec difference between the right and left delay to prevent image problems caused by the Hass effect. To manufacture the above scenario, offset the left and right delays from the actual distance delay by at least 15msec and no longer than 80msec. To create dimension, have the left delay arrive 30 msec and the right delay at 45msec. Theoretically the right delay (45msec) should sound slightly lower in level and in high frequency content but for the purposes of creating dimension this is unnecessary for you most likely won't be listening to only one signal in a production and be in a position to detect the location of the delay (reflection). If the delay times are quite different in level and in time with each other (between 15msec and 80msec) you will notice an effect like you are sitting close to a wall while listening to the sound source.

E.g. A singer panned dead centre, a delay of 20msec at -3db panned hard left, a delay of 80msec at -6db, will give the impression to the listener that singer is some distance at the left side of the image and the listener is sitting close to a wall on the left side.

If the want the listening position to appear further back, have the left delay at 75ms and the right delay at 60 ms, both at a lower level and even less high frequency content. Even though there is a difference in arrival times of the left and right delays (15msec), the effect of dimension will greatly over-ride the slightly off centre listening position if the direct sound is panned in the middle. It is through this extra delay and altered frequency response that contributes depth to the direct sound. We must also note that the type of envelope, a percussive attack or a slow attack will determine the delay time as sounding dimensional or discrete. The frequency response of the delay will determine the absorption coefficients of the reflective surfaces. What occurs, the psycho-aural response is alerted, which tells you that you are listening to the sound at a distance in a reflective environment. Where as if you just heard the original sound only without reflections, the psycho-aural response would suggest you are listening to a signal while standing elevated in the middle of a field. If you had a signal panned centered and duller sounding reflections of 40ms (left) and 60ms (right) it would sound like you were sitting at a distance, slightly left of center to the left, for the left reflected delay is slightly closer, brighter and louder in relation to the right reflection.

If these reflected signals are very dull sounding, it will imply that the reflective surfaces are absorbing the high frequency content and placing you in an environment of wooden walls rather than glass. When a signal bounces off a surface it will always sound duller than the original for any type of surface absorbs some sound. The duller the reflection, the higher the absorption co-efficient of the reflective surface. If a reflection is heard after 100-150ms (approx), you will perceive it as a separate form of sound energy and as a discrete delay. When the delay is discrete, it will be easy to localize in the stereo image and might prove to be distracting. So if you have a reflection coming in at 200ms, and it's panned to the left side, you will hear it coming directly from the left and will not prove to be beneficial in creating depth for it is detached from the original signal. For example if you had a percussive instrument like a snare drum and you wanted to add depth to the sound, the delays will have to be in the vicinity of 15-60ms. If the delay is any longer it will sound discrete, for you now hear the difference between the transient of the original drum and the transient of the delay resulting in a confusing sound. A good rule is to remember for adding dimension with percussive elements, the faster the attack of the sound envelope, the shorter the delay will have to be to prevent a discrete delay from appearing. If you would like to create a slap effect coming from the rear to simulate a canyon, then go ahead and add in discrete delays but make sure they are at a lesser volume, duller and not at a time setting that is also a rhythmic factor in the tempo of the piece of music, for the delay will most likely land on a half, quarter, eight or sixteenth note of the tempo and will be masked and hard to hear as a dimensional contribution to the sound.

4) If the instrument happens to be a piano or guitar playing with even dynamics, the delays can be approximatley15-100ms. If the instrument is a violin, the delays can be 70-120ms. The slower the attack, the longer the delay can be in achieving dimension. In a surround setting, if you add in additional longer non-discrete delays to the rear channels you will create an even more realistic listening environment. The delays in the rear channels will have to be longer than the 2 delays in the front left and right channels, yet short enough that they don't sound discrete in the rear channels. Another good rule is when adding in longer delay times, dampen the high frequency content of the delay as the time gets longer. This will create the illusion that the signal is losing fidelity because it is traveling over a longer distance than the original and also indicates that the reflective surface is further away from the listener. Another situation to factor in is that the duller the delay is, the higher the reverb co-efficient of the reflective surface. A delay's frequency response can therefore dictate the reflective properties and distances of the reflective surfaces of the listening environment. It is up to the engineer's discretion on how they want to manipulate the sound of the delays to simulate a realistic listening situation. This creative manipulation of delays works very well with audio that have slow to medium attack times with harmonic content, ambient sound and effects. Generally, any reflections arriving between approx15-100ms in the front channels and 50-150 ms in the rear channels will not affect clarity when equalized and mixed in accordingly. Adding reverb with these delays will create a natural sounding acoustic environment. With additional reverb and correct pre-delay settings, one can create a more realistic environment.

   NB: The individual level and frequency response of these elements determine the size and the quality of the listening environment.

In figure 11, you will see the layout of a concert hall with different sound location sources situated at fixed distances from the optimum visual and listening position. The goal here is to figure out what elements contribute to the overall sound from the 3 different listening positions-seated centre but at differing distances from the sound source. If one can figure out what is involved in what is happening to the audio signal at these 3 different listening positions, then it would make sense that if we reverse the scenario where the listening position is stationary and the sound sources can be placed at different distances, than one should be able to manipulate the audio elements to create dimension at the listening position in the theatre. Instead of the listener having to physically move to hear 3 different perspectives, the sound source can be moved to different positions and depth.

Audio Reflections

The best sounding mixes in surround have dimension and perspective where one can actually visualize depth in the music. To achieve this one needs to understand how direct-sound, reflected- sound and reverb work with each other. How to alter these elements when you are mixing to achieve desired dimensional perspective in creating dimension. Dimension is simply a combination of multiple delays (reflections) and original sound.

Once reflections get dense enough, that you can no longer distinguish them as separate individual sounds, they turn into diffused reverb. To use depth effectively one needs to look at music sounding 3-dimensional rather than a 2–dimensional. With creative use of these elements, level, frequency response and time duration you will have the basic knowledge on how to create dimension in mixing. However there are fundamental laws of physic that need to be adhered to when trying to create believable dimension. If you are into creating dimensional landscapes one should posses a basic understanding in how human hearing relates to audio and how to manipulate the various elements. As they say; "If you want to break the rules, you need to know the rules you are breaking". In this age of digital technology, artificial reverberations such as convolution reverb algorithms* are not only more affordable than ever before, but can be easily manipulated in creating believable realism.

With a good understanding of the physics of natural acoustic environments, and the fundamental operational principles of reverb processors, it is possible to quickly create the illusion of any acoustic environment you can imagine. First, one needs to know how sound arrives to the ear in certain listening positions in a concert hall and how to re-create this listening position if you want realism in your mix.

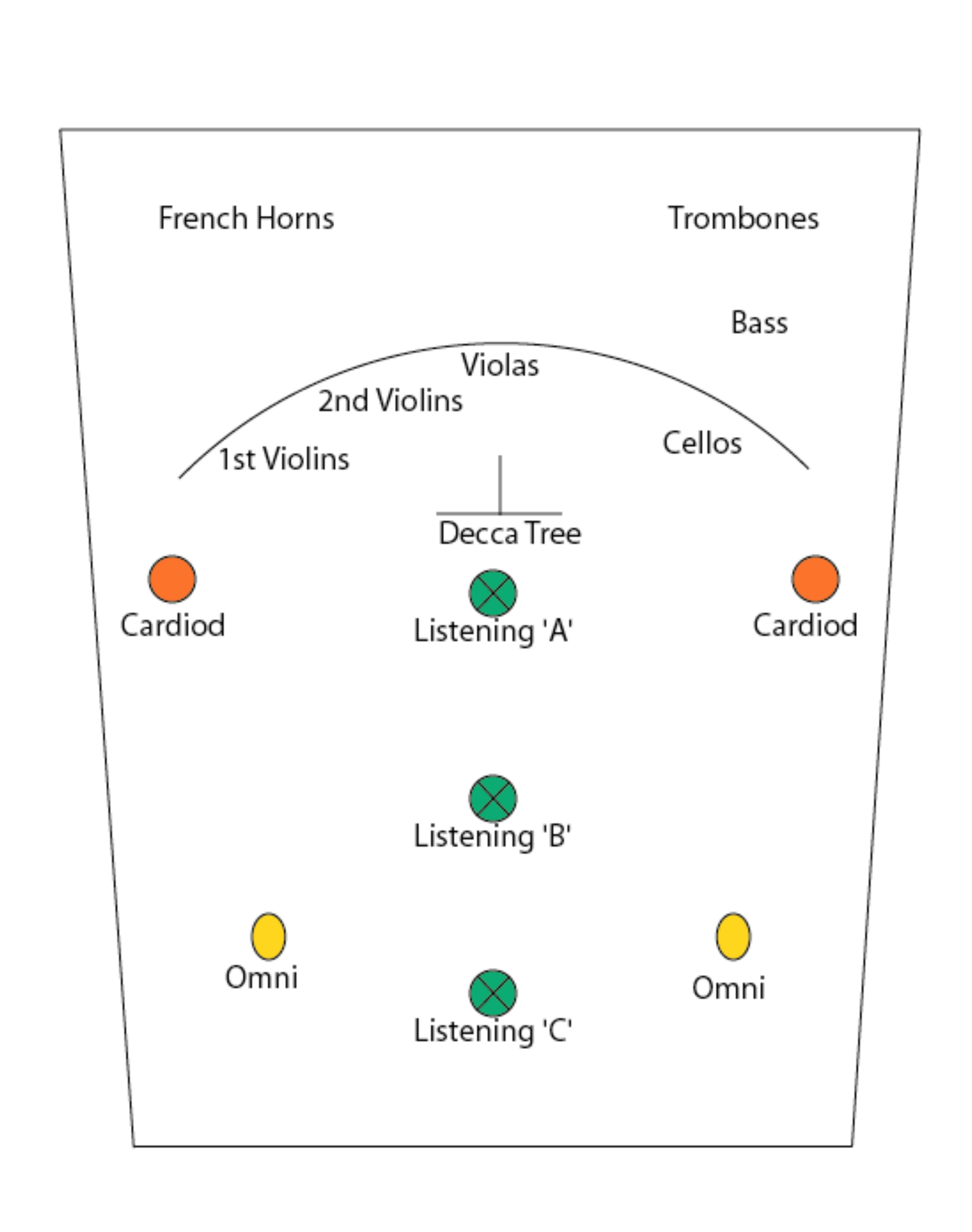


Figure 11: Three listening positions in a concert hall

Realistic Surround-Creative Surround

The listener will hear amplitude in the following breakdown, the percentages are approximate in order to define the concept;

Listening Position ' A'

85% Direct Path

1% Early Reflections

14% Reverb

Listening Position ' B'

70% Direct Path

15% Early Reflections

15% Reverb

Listening Position ' C'

54% Direct Path

23 % Early Reflections

23% Reverb

NB: These ratios are approximate and are used to distinguish the different levels of the 3 elements to illustrate dimension. In all cases the direct path amplitude will always be the greatest unless impeded by a physical structure, in which case the early reflections and reverb will contribute 100 % of the total sound.

Listening Postition 'D' (N0 Direct Path)

0 % Direct path

50 % Early Reflections

50 % Reverb

NB: In position 'D' the amplitude relationship will change according to the listening position's location to the walls, floor, and ceiling. If the listening position is equal distance from all the walls, floor and ceiling the reverb content will make up most of the total amplitude especially if all the surfaces are not perpendicular or parallel to each other. As we will discover this type of sound design strategy can play a role in creating dimension outside of a standard mix.

Placement 'A'

In placement "A" the original direct sound (85%) will be full-frequency response and arrive to the listening position in approx 1msec (one-meter). The early reflections from the front and surfaces of the hall will be very low in level compared to the direct sound because the listener is sitting very close to the original sound source and far from the walls therefore the level of the direct sound will be substantially louder than the reflections. For all intents and purposes, we will look at early reflections as being unnecessary in the overall sound.

In position "A" the reverb will be delayed when it arrives back to the listening position for it takes a significant amount of time for the audio signal to reach all the surfaces and diffuse itself into reverb before it arrives back to the listeners ear. The frequency response of the reverb will show that the mid-high frequency content has been rolled off. The actual time difference between the onset of the reverb signal and the original signal and will be approximately between 100-150ms depending on the size of the hall and the environment you are trying to create.

NB:

The frequency range of the reverb dictates what type of material is being used on all the surfaces (absorption coefficients). The softer and rougher the surfaces are, the duller the reverb. If the surfaces are made of wood, the reverb will sound warm and not contain a lot of high frequencies. If the surfaces are concrete and glass, the brighter the reverb will be with a longer RT-60 time. As reverb time decays it's high frequency content in the overall reverb amplitude decreases. Over distance and time, the atmosphere eats up high frequencies and the reverb diffusion increases as the reflections keep bouncing from surface to surface. In other words, as the reverb decays so does its high end no matter what type of surface it is.

When sitting in a hall in the A position, the listener will hear the audio signal very clear with a full frequency response, almost no early reflections and a warm reverb arriving approximately greater than 100msec.

When trying to recreate this image in surround sound mixing, you would have to make sure the signal is very present sounding and if you feel the need to enhance the low and high frequencies to make the source sound more intimate then do so. The reverb should be warm sounding so you will have to factor in a high frequency roll off on the reverb return to give you contrast between the original sound and the reverb. If the reverb is as bright as the original, it will confuse the ear, for you would be creating a scenario that in reality would not exist or the reverb environment you are to trying to emulate would be an appalling one.

If you wanted a lead vocal to sound like it is very close in front of you and in a beautiful warm sounding environment, then try this approach:

1) Enhance the high frequency range between 12khz-16khz.
2) If compression is required, use it as dynamic management and very transparent.
3) Decide on how reflective the environment should be-a good starting point is 2.5sec-3.5sec.
4) Decide on what type of surfaces are in the environment you want to emulate. Factor in a high frequency roll off with a slow smooth curve. I would suggest you go down as far as 3khz as the -3db point. Remember to have the space in the track where you will be able to appreciate the reverb quality. If there is a lot going on in the mix, try a higher roll off and a shorter reverb time.

5) Decide on how far the reflective surfaces are from the sound source and back to the listening position. This will be your pre-delay setting. The longer the pre-delay the further the surfaces are away from the sound source and the listener. A good starting place is a pre-delay time between 100msec and 150msec.
6) With vocals there is sibilance and all artificial reverb units do not have an algorithm to solve this dilemma. A reverb environment with a lot of sibilance is unrealistic and simply dreadful sounding. You lose presence from the original signal for the reverb sibilance contains a lot of the mid-high frequency, which ruins the overall effect you are trying to create. Take the vocal and insert a desser over the reverb send-not over the actual vocal. You can also assign the vocal to another channel (post fade from the first channel) and insert a desser and send to the reverb from the channel. Only use the direct sound from the original channel!
7) If you have the luxury of an open sounding production such as a ballad you can add in delays to the reverb sound. This is not to create a realistic sounding environment, but is used for creative purposes. The goal here is to extend the harmonic component of the original vocal so it can sound even more melodic. What you need to do is first figure out what the tempo of the song is. If it is 100bpm, then the quarter note delay is 600msec. Adding this delay into the reverb sound extends the melody in the overall reverb sound. The goal is to add the delay into the reverb where it is not noticeable and this is why you need to pick a delay time setting that is rhythmically related to the tempo of the song. If the delay is a fundamental of the tempo it will land on rhythmic grid that relates to the rhythm of the song. If the delay was 550msec and you wanted to add the delay to the reverb, you wouldn't be able to add in that much level of the delay for you would hear it sounding ahead of the beat which would prove to be distracting. If the delay is a fundamental of the rhythm then the delay will land in tempo to the song and will be masked, for there more than likely be an instrument playing on the same beat. Not only that, the A-section of the delay waveform will be hard to hear because it's landing on the beat, which is a good thing, for as previously stated the A-section contains mostly mid-high frequency and little sustain or music. What a bargain I say! Don't forget to incorporate a high pass filter on the delay because we know that delays have to sound duller than the original in order to sound convincing. Also insert a de-esser over the delay send and do not be afraid to de-es heavily and all frequencies above 2khz. This will get rid of the component of the signal that is basically nothing but noise, and reverb should be noise free. When you have a chance, listen to some of your favourite recordings with the de-essing thing in mind, and you might discover yourself saying, "It's a good sounding track, but I think it would have sounded better if they got rid of all that sibilance-noise in the reverb." With the option to assign four delays to all the surround speakers to enhance the melody-

music in the reverb you will use a delay setting that is rhythmically agreeable with song. However you can't assign 600msec to all four channels for the delays will all collapse into mono in a position that is dead centre in the room. What does work is an alteration of the delay setting. Try this setting: front left 592msec, front right 608msec, left rear 577msec, right rear 623msec. You will notice that all the delay settings are centered on the 600msec-quarter note setting. The delays are far enough apart from each other to avoid imaging or phasing problems, yet sound distinguishable from each other. Add in a little regeneration and roll off and bring the level of the delays up to a level where they enhance your reverb and away you go. Do not be afraid to bring up the level of the delays to a point where they do sound quite noticeable with reverb, because the delays will be landing on a fundamental of the rhythm and the A-section of the delay will be masked.

Here is an example a template for position 'A'

Reverb Pre-delay 80-150ms (RT=2.5-3.5 sec)

High Pass 2.5khz-5khz

Here is an example a template for position 'A' (reverb enhancement)

BPM=100=600msec for quarter note

Front Left 592msec (less high freq than original)

Front Right 608msec (less high freq than original)

Rear Left 577mecs (lesser high freq than FL delay)

Rear Right 623ms (lesser high freq than FR delay)

Reverb Pre-delay 80-150ms (RT=2.5-3.5 sec)

High Pass 2.5khz-5khz

Delay Regeneration=Reverb Decay time (RT=2.5-3.5 sec)

NB: Make sure that the rear channels of reverb and delays are as audible as the front channels, if the listening position is closer to the front channels than the rear channels.

In position ‘A’ you are trying to place a sound source 15 feet in front of you with the goal of the source sounding close and intimate. The original signal needs to reflect off the walls for a while to create reverb and then make its way back to the ear, yet sound distinct from the early reflections. The time for the reverb to arrive at the listening position has to be greater than the time of the latest early reflection for it to make sense and sound believable. The frequency response of the reverb will depend on the reflective properties of the walls. If you wanted the hall environment to sound warm you will have to incorporate a high frequency roll-off on the reverb return. Because the ‘A’ listening placement is not close to a wall and at a distance to the original sound source you will barely hear any early reflections. Adding in delays to the rear channels the delay of the onset of the reverb will indicate how far the walls are from the listener. The length of the reverb will indicate how live the environment is. The overall sound will be intimate, clear, and pleasing to the ear especially if it is a great singer or soloist performing a ballad. To create this in mixing you will need to add in a reverb that rolls off more high frequency content over the decay of the reverb which means as a reverb gets longer it also gets duller. Roll off the reverb return in the high frequency and low frequency area and maybe slightly boost around 2-2.5K to add a little presence for clarity in the reverb. Watch out for low frequency build up that might clutter the mix. Incorporating a low frequency roll-off in the reverb return around 150hz will help in maintaining articulation in the reverb. In total the original sound will be 80%, early reflections 5% and reverb 15%.

Listening Position 'B'

In listening placement 'B' (the exact middle position of the theatre) the original sound source (70%), as in position 'A', will have slightly less mid-high frequency content due to the increased distance between the sound source and the listener. The early reflections (15%) arriving from all the walls will be louder in relationship to the original direct path source. As earlier stated, the larger the time difference between the arrival of the direct path source and the arrival of the early reflections between 15msec and 80msec will dictate that the listening position is further away from the sound source. Delays (from the front) of 40 and 55ms will indicate that the listener is sitting further away from the sound source than with a delay of 20 and 35ms. The difference in amplitude between the direct sound source and the early reflections will decrease as the reflections increase in time difference between the direct sound and the early reflections. In other words, if you physically moved your listening position back a couple of rows from centre position 'B', the amplitude of the reflections (and reverb) increase and the amplitude of the direct sound decreases. If for example you were standing in the foyer of a concert hall the level of the direct path, reflections and reverb would almost be identical. When the early reflection times are fixed, the amount of level of the reflections will always be lower than the direct path source. If the early reflections are low in amplitude, it would suggest that there are no reflective surface, or if there are reflective surfaces, they are absorbing a lot of the sound energy.

To construct dimension for the 'B' position, the early reflections from all four-surround channels should arrive close to each other in time and the same amplitude. Let us pick the dimension time between 15msec and 75msec, '45msec'.

For the front channels, FL-38msec, FR-31msec, RL-52msec, RR-59msec.

This set up will create the illusion that the back wall is further away from the front wall. To compensate for this, in a way that will work, increase the levels of the rear delays or switch up the delay times between the channels. What we are trying to do here is create dimension, so breaking some rules need to be executed in order to create the dimension that is required. Another idea is to insert voltage controlled oscillators across the delays, but the rate and depth would have to be minimal to avoid pitch errors.

The reverb will arrive to the 'B' listening placement closer to the original signal than in the 'A' listening placement. This is because in the 'B' listening position, the time difference between the direct path source and the early reflections (and reverb) decreases. The early reflections (and reverb) frequency response will sound closer in relation to the original signal due to the slight degradation of the original sound over distance. In an ideal situation of a middle position in a concert hall, the front reflections and the rear reflections So to create this dimensional effect make sure that the original sound source does not have an extremely wide frequency response. The depth will be created by two or delays arriving to the listening position between 40-55ms. In the rear channels add in delays of 70-85ms to create the illusion of rear reflections. Make sure all delays are equal in level have some high frequency roll off so the ear will not confuse the delayed signal with the original signal as being the focus. As previously stated the reverb pre-delay will have a smaller pre delay time than position 'A' and sound slightly brighter to the slight degradation in the 'B' original signal. The purpose of the 'B' placement allows you to add depth and perspective to audio that needs to situated in a placement that doesn't fight with audio in the 'A's placement.

Here is an example a template for position 'B'

Front Left	35m	(less high freq than original)
Front Right	50ms	(less high freq than original)
Rear Left	80ms	(lesser high freq than FL delay)
Rear right	65ms	(lesser high freq than FR delay)
Reverb	Pre-Delay 75msec-	(RT= 2.0-3.0sec)
High Pass	4.5khz-7.5khz	

Listening Position 'C'

In placement position "C" the direct path sound, coming from the front (54%), will arrive to the optimum listening placement at a lower level than positions "A" and "B" and its frequency bandwidth will be even less than listening placements "A" and "B". The early reflections (23%) will be even longer and louder and the reverb (23%) decay time will remain the same. It is interesting to note that the difference in arrival times of the direct path sound and the early reflections and reverb will be smaller, creating the illusion that the sound source is more distant sounding. If one were to analyze the duration of the wavelength of listening position 'A' and listening position 'C', you would notice the episodic time value of position 'A' would be longer than position 'C'. By episodic sound, I am referring to when one first hears sound until the end of the decay time. As the listener moves further from the sound source the three elements that make up the total sound move closer together in arrival times. The sound quality in respect to full frequency bandwidth also minimizes between the different elements as the listening position moves further from the sound source. If one was in the listening position 'A', the difference in sound quality between the direct path sound and the early reflections/reverb would be more noticeable than the 'B' and 'C' listening positions.

With mixing for the 'C' position one should not go out of your way to deteriorate the sonic quality of the direct path sound. The idea is not to enhance the frequency spectrum of the direct path sound but to leave it more natural sounding or even slightly duller in equalization, especially in the range above 12khz. It should not contain a lot low frequency information for that would just muddy up the sound. It should contain mid-range presence in keeping with the character of the instrument in establishing its rhythmic and harmonic content.

The early reflections to be created will be at the end of the depth spectrum of delay settings, before the delays sound discrete from the direct path sound. The reflections sound quality will sound closer to the direct path sound, whereby the equalization contrast between the two will be minimal. For the front channels, create delays from 55-70msec in to add in the illusion of front and side reflections. With the back reflections, try 75-90msec, being careful with transient sounds that might start sounding discrete.

The reverb will also contribute more to the overall sound and its pre delay time will be even shorter in relation to the direct sound originating in the 'C' placement. Because the acoustics of the environment are fixed, the reverb decay time should not change dramatically, and if anything get shorter than the reverb time of position ‘A’, for the arrival of the direct sound and the reverb will be very close together. With Reverb to be believable for placement “C”, the pre-delay needs to be small. When the pre-delay gets smaller in time, it indicates to the listening placement that there is distance.

Here is an example a template for position ‘C’

Front Left	50msec	(a little less high freq than original)
Front Right	65msec	(a little less high freq than original)
Rear Left	80msec	(lesser high freq than FL delay)
Rear right	95msec	(lesser high freq than FR delay)
Reverb	Pre-Delay 20-30msec-	(RT= 1.5-2.5sec)
High Pass	7.5khz-10khz	

Creative Placement

When creating dimension that is realistic, one should also factor in extending the boundaries of the fixed listening environment. If you wanted a sound source to appear form the rear channels all one would have to do is pan the source there and factor in the delay and reverb settings form the previously mentioned listening positions. However if one wanted to define depth and location, one should locate the early reflections/reverb from the same source as the direct path sound now matter where it was originating.

If you wanted a marching drummer to appear that it is moving from a distance behind you but getting closer, you would have to create delays and reverb with the above settings to suggest what position you are trying to create. In mixing one could pan the direct path dead center in the rear channels and add in delays that will dictate the distance and a reverb setting that will define how large and how reflective the environment is. By changing the balance of the individual amplitude settings between the direct sound and the early reflections/reverb, one could create the illusion that the marching drummer is at some considerable distance behind the listener and moving closer to the listener. This effect could also be created for images with changing depth coming from the corners in surround by creating a stereo spread across the mid-left side and the front center side; where the image would be appearing from the front left location at a distance and moving in closer. This could be utilized across any two locations by creating discrete stereo dimension. If you have the space in the mix that one could easily hear these ideas, you should take advantage of it in order to add more dimensions in your mixing.

Time Panning

If the sound source is not directly in front of the listener, it will be closer to one ear (ipsilateral ear) and therefore arrive at this ear first, leading to the ITD cue. In the artificial stereo technique, this cue can be simulated by simply sending to the contralateral ear, a delayed version of the signal sent to the ipsilateral ear. For example, when the desired sound source position is to the left of the listener of a stereo setup, the right ear will receive a delayed version of the signal sent to the left ear. The amount of delay determines the position of the virtual source and therefore, by allowing for a variable time delay, the virtual source may be positioned between the two loudspeakers. The time delay actually required to position the sound source to either the left or right loudspeaker is rather small. Experiments indicate a delay of between 0.8ms to 1.4ms. Given this short range of delays required to position the virtual source to either of the loudspeakers, the effect produced by this technique quickly degrades with even small listener movements, especially side-to-side movements. Movements of a few feet may lead to time delay’s which are much greater than the small amount described above. In other words, the extent of the sweet spot is very small.