

Basic Principles of Audio Design in Conferencing

by Ryan Shannon 2014



Participating in a video or audio conference is an individually subjective experience (indeed, every experience is individually subjective). This idea is the ultimate defining truth in the AV business. As an AV sales person, design engineer, or consultant, you know that the end users determine the success of any conference system. If the end users are happy and the conferencing experience is natural and seamless, you are in fact, successful.

The challenging duty of the sales person, engineer, or consultant, is to use scientific knowledge and previous practice to try to create a pleasant experience for all end users. There is as much science involved as understanding contemporary human behavior and perception. Understanding the needs of the end users and the demographic diversity is crucial to the success of the system design. The audio portion of the goal should be relatively straightforward; to provide natural sounding speech that is clean and clear, ample in volume, and greatly intelligible. To obtain that, there needs to be a substantial understanding of acoustics, microphones, and best practices. We will examine various microphone types and a variety of approaches to bring forth their strengths and weaknesses. In the end, I will make a strong case for leading with specific microphone types and design approaches, highlighting the simple but overpowering benefits that make it easier to obtain highly intelligible audio for your end users. Producing high quality spoken word audio should be relatively easy if you understand some of the most basic audio truths. The first stage in the audio path comes from acoustic energy, so let's start by simplifying acoustics and its most important concepts regarding intelligibility.

"Acoustics refers to the study of sound, namely, its general transmission through solid and fluid media, and any other phenomenon engendered by its propagation through media. Sound may be described as the passage of pressure fluctuations through an elastic medium as the result of a vibrational impetus imparted to that medium" [Raichel]. Let's simplify acoustics and

uncover the basics that effect intelligibility in conferencing. There are three phenomena that occur when dealing with sound: Reflection, absorption, and transmission. Sound is either reflected, absorbed, transmitted, or most often a combination of the three. Reflection and absorption are self-explanatory. Transmission is when sound leaves or enters a space, similar to travelling through a window or passing through a wall.

When sound energy comes into contact with a dense and reflective surface like marble, a small portion of that sound is absorbed by the marble and converts into heat. Also, a small portion is transmitted through the marble. Most of the sound will be reflected off the marble with a relatively small amount of energy loss. The opposite occurs with a highly absorptive material; most of the sound is absorbed and converted into heat, a small amount is transmitted, and a small amount is reflected at low energy levels. So when we survey sites that are getting conferencing systems, keep these things in mind to determine how reflective or absorptive a room will be in relation to sound energy. And in terms of transmission, observe how the exterior elements are affecting the room. Is there a busy street right outside the room causing traffic noise to be transmitted through the windows or walls? If so, that will cause the ambient noise level to increase. The exterior elements and materials in the room that make-up the surface area will have a large impact on natural and reinforced intelligibility.

Inverse Square Law

The inverse square law states that as sound travels, the sound pressure level is reduced by a factor of four as the distance doubles, as illustrated in figure 1. This is true when the sound source is well away from any surfaces that might reflect the propagating sound wave. This means that when a person speaks in a conference room, the sound pressure level of that speech will be close to four times lower at 6ft than it will be at 3ft. This concept proves the reasoning behind microphone type and placement being extremely important to achieving strong speech levels.

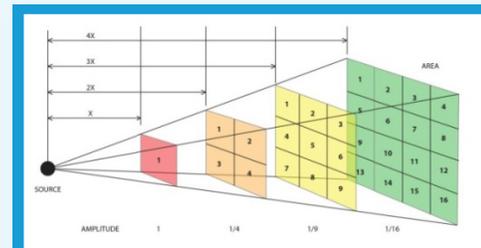


Figure 1: The Inverse Square Law

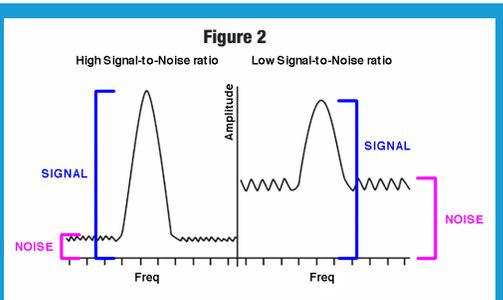
For spoken word applications like conferencing, a good rule of thumb for microphone placement is to be within an arm's reach distance from the microphone. The best-case scenario is to use a lavalier or wearable type microphone that is affixed to the source so that the distance is minuscule and never changes even if the source moves locations.

Signal to Noise Ratio (SNR)

When noise level amplitudes are produced in close proximity to speech level amplitudes, intelligibility decreases due to an acoustic phenomenon called sound masking. In figure 2, the image on the left represents good signal to noise ratio. The signal level is very strong and noise level is very low. The image on the right represents relatively less than good signal to noise ratio. The signal level and noise levels have less separation between them. Signal to noise ratio is probably the most important concept to keep in mind for reproducing highly intelligible speech.

Intelligibility is largely determined by the amplitude separation of the unwanted noise of a space and the desired speech levels.

Maximizing the SNR is absolutely imperative to yielding a pleasant experience for the end users. You must put a design in place that puts SNR at the forefront. If the ambient noise level of a space is 20dB SPL and your microphones are capturing speech levels at 75 dB SPL, you have a great deal of separation between the two amplitude levels and the speech will be highly intelligible.



The plot on the left represents good signal to noise ratio. The signal level is very strong and noise level is very low. The plot on the right represents a poor signal to noise ratio.

at 55dB SPL, intelligibility will suffer. For example, have you ever had trouble carrying on a conversation in a loud restaurant? The intelligibility is low in that environment due to the noise levels existing at similar levels to the natural speech levels. This is especially common in restaurants that have many dense, reflective surfaces and lack absorptive materials. The sound energy lingers in these environments because it is continuously reflected rather than absorbed. Also, people will speak loudly to overcome ambient noise levels, and by doing so they help keep the ambient noise levels high in the space. You have to speak loudly in that environment in order to be intelligible. What you're doing is creating separation between

the speech level and the noise level. It's also worth mentioning the "cocktail party effect", which refers to the ability to focus one's listening attention on a single talker amongst a cacophony of conversations and background noise [Arons].

In conferencing, the concepts of the inverse square law and signal to noise ratio hold accountable the success of the design. You must capture strong speech levels by controlling microphone placement proximity, especially in spaces where the ambient noise level is higher than normal.

Humans possess the ability to focus their listening on a particular voice when in a high ambient noise environment. This is relevant in room acoustics topics but does not apply to our topic of audio in conferencing. In most cases, the cocktail party effect will not help one's ability to increase intelligibility in conferencing because of the lack of natural acoustics and binaural cues as well as the lack of clear lip-reading.

Microphones

Microphones essentially perform the same function as ears. They have diaphragms that are displaced by air molecule movement, converting sound pressure level into electrical energy. Microphones have pickup patterns, which refer to the direction(s) in which the microphone will best capture sound. Directional microphones capture sound from the front and reject sound from the rear. Figure 3 is an example of a polar chart, which illustrates the pickup pattern of a particular microphone. The green line indicates the level of sensitivity relative to the direction of the sound source. The reason that directional microphones exist is so that the user is able to focus the directionality of the microphone at the desired source while the microphone rejects sound from other directions. This is done to keep separation of the desired sound and the unwanted sound. Some environments have ambient noise levels as high as 55dB SPL. This creates an extremely challenging environment for maximizing SNR and intelligibility. Capturing high speech levels is most easily done by getting the microphone close to the talker.

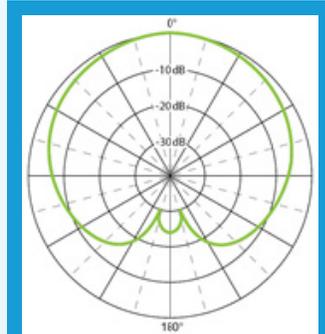


Figure 3 - Polar Chart of a Directional Microphone (Cardioid Pickup Pattern).

Choosing a wearable or gooseneck microphone will be your best bet to overcome such a high ambient noise level. Capture speech levels as high as possible in order to create separation from the high ambient noise, maximizing your signal to noise ratio thus yielding better intelligibility. ➤

Noise builds up easily and quickly, especially in spaces where there are a lot of dense and reflective surfaces. In recent years, we have seen minimalism and clean lines become popular approaches with respect to interior design. That means there is less “stuff” in the space and on the walls that would otherwise help to absorb sound. Today, when you look at new construction in high budget commercial spaces, you will find a lot of reflective surfaces like floor to ceiling windows, large and dense tabletop surfaces, and stone countertops. All that surface area that is highly reflective helps reflect and reinforce the unwanted noise of the room. These spaces are referred to as “live” rooms because they are easily excited or energized by sound energy.

In this type of space, sound will linger longer than an absorptive room because the energy continues to reflect rather than absorb. This is a harsh acoustic environment for speech intelligibility and is very common in conferencing spaces today. Poor microphone selection and/or placement in these spaces will annihilate intelligibility.

Issues with Common Practices

Let's explore ceiling microphones and use them as our example of poor design solutions based on the laws of physics. The advantage of a ceiling microphone is that it is off the table and out of the way. Aesthetically, that's a great thing. Unfortunately, being close to the ceiling also means ceiling microphones are closer to common noise sources such as HVAC vents and returns, projector fans, and light ballasts. The microphones will have a better chance of picking up these noises because of their proximity to them, reference the inverse square law. This noise then enters into the audio signal with the speech, which decreases the signal to noise ratio and decreases intelligibility due to sound masking. Ceiling microphones are a recipe for a subjectively inferior experience. They can be a costly experiment, as often the end user will eventually be left with no other option but to replace them with a different solution. In a “dead” space which is highly absorptive and has very low ambient noise levels, ceiling microphones may work fine. However, the minute that unwanted noise builds up and the users begin to speak softly or quickly after one another, the ceiling microphone design solution will fall short of ideal.

Ceiling microphones are commonly omni-directional. They are picking up sound and noise in 360 degrees. A talented audio engineer can minimize ambient noise by using various advanced techniques like external muting, having a lot of experience in tuning noise cancellation, and by using noise compression and gating. However, the end result will never be as good as a microphone design solution that puts the microphone closer to the talker. With a closer proximity to the talker, the speech levels will be maximized creating that separation to the ambient noise levels. If that same talented audio engineer can make the ceiling microphones sound good in most cases, imagine how much better a wearable or a directional tabletop microphone would sound with much less work involved.

The audio engineer will not have to apply as much processing to a wearable or tabletop microphone. The speech will sound cleaner, louder, more natural, and more intelligible which covers all of the audio goals. The simple truth is this; a ceiling microphone will never sound as good as a wearable or a tabletop microphone, simply based on the proximity of the microphone to the source. Speech levels will naturally be captured at higher levels with a closer proximity, supported by the inverse square law.

Deeper into Physics of Conferencing Audio

Let's go a little deeper into the physics of sound as it uncovers why ceiling microphones are a less than ideal solution in conferencing audio.

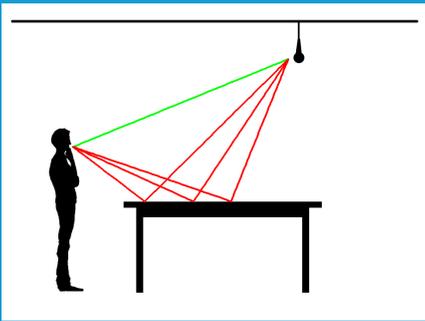
Level Strength

In our common application, the boardroom, the table and the talkers' bodies will reflect the speech defining frequencies forward, up, and away from the table. So, because there are surfaces reflecting the propagating sound, we may not have 6dB of loss as the distance doubles. We will likely have 4-5dB loss as the distance doubles.

Now if a tabletop microphone is placed two feet from the talker and a ceiling microphone is six feet from the talker, your distance between the talker and the two microphones has tripled. The talker is at zero feet, the tabletop microphone is at two feet, and the ceiling microphone is six feet. So let's say that the tabletop microphone is picking up 60dB-SPL. At four feet, a microphone would pick up 55-56 dB-SPL. And at six feet, the ceiling microphone will pick up 50-52 dB-SPL. This means that there will be an 8-10dB difference between the two microphones' input signals. To help understand what this will be perceived as, a 10dB change in sound pressure level is perceived as twice or half the volume. So the pre-processed audio from a ceiling microphone will be perceived as half as loud as that of a wearable or tabletop microphone.

Reflected vs Direct Sound

End users will likely not be looking up at the ceiling microphones as they are speaking. This means that much of the sound received by the ceiling microphone will be reflected sound, which will have less energy than the direct sound because a portion of it will have been converted into heat, and also a small portion lost from transmission ➤



Much of the sound received by a ceiling microphone will be reflected sound.

through the materials that caused the reflection. Signal to noise ratio will not be maximized because the signal will be composed predominantly of reflected energy. Reflections off the table, the users' bodies, and other surfaces in the room, will arrive

at the microphone at varying times and some at similar levels. This arrival time difference between all the reflections and the direct sound can cause the blending of consonants. Subjectively, this experience will yield lower intelligibility, cause distraction, and can lead to a decrease in interest and focus on the content.

Noise Level

Ceiling microphones are directed down at the end users where papers are being rustled and notes are being typed or written. The microphones will likely capture that noise at similar level as the speech, which affects the SNR and intelligibility. Also, without having the capability to press a mute button on the ceiling microphone, there may be more microphones capturing sound than needed. What if the room is only half or a quarter full? You may have more microphones than needed capturing reflecting sound and unwanted noise. This will potentially introduce a lot of unwanted noise into the signal. With wearable or tabletop microphones, there's the option to press the mute button on the microphone, which is placed directly in front of the user. The users will only use the number of microphones needed, which will help to keep the noise levels low.

Digital Signal Processor (DSP)

Digital signal processors are used as the source of the corrective audio processing. They are designed to mix and send sources to their proper destinations, as well as provide acoustic echo cancellation and other advanced sound processing elements. DSPs are vital to the successful tuning of any modern conferencing room and system. The simple truth to take away is this: the ideal sound design scenario is that the microphone design yields the best signal to noise ratio and strongest speech level possible prior to the signal entering the DSP. If the audio is the cleanest and strongest signal possible prior to entering the DSP, any corrective processing that is needed will be fairly simple and minimal. This puts the sound engineer in the best position at the start, which is to have many tools to use in order to make a good signal sound great. Applying the least amount of processing needed will ensure that the speech stays natural sounding. Applying a

lot of processing to speech signals can often make the speech sound unnatural. Unnatural sounding speech will be a distraction to the end users which can render the experience unpleasant and thus unproductive if interest and focus is lost. Remember, the success of a conference system design boils down to the subjective experience of the end users. The end users will perceive your design to be successful if there are no perceived distractions and negative feelings based around intelligibility.

End User Behavior

Most end users are not knowledgeable when it comes to the proper techniques of using microphones. That is why it is so important to consider every factor relative to the application and the space. Intuition and attention to detail really separate the great sound designers from the average ones. In order to achieve our goal of providing a subjectively pleasant experience for all, you really have to put together the best design solution possible to combat all of the acoustic challenges of the space.

Understanding your end users and their behavior habits should be taken under serious consideration. There are so many factors and variables that will affect the subjective experience of a conference, and that is why it is imperative to make an effort to only recommend the best solution available. As an AV professional, we are paid for our expertise in these matters. We have to be able to predict good and bad outcomes due to end user behavior. We have to assume that the end users will likely do the things that will hurt intelligibility, like leaning back in their chairs and looking away from a microphone while speaking. Or, cracking open a soda can that is placed right in front of a microphone.



Microphones which have a mute facility make for a much better end user experience.

Training and sharing the simple audio truths with end users can help assure your design's success. Provide laminated quick guides and show the end users the proper microphone placement in order to attain great signal to noise ratios. Use the rule of thumb for tabletop microphones; the furthest you should be from the microphone is an arm's length.

Explain the idea of keeping the speech levels high and the noise levels low. Urge them to use their orator's voice and speak from their diaphragms. The educated end users will spread the knowledge to other end users that are mis-using the technology, and correct behavior habits will spread. The more end users you can educate, the more successful your design will become. ▶▶

Recommended Microphone Types

If ambient room noise is not terribly high, tabletop microphones are a good choice for the design because the end users don't have to wear the microphones. As a tabletop microphone option, a gooseneck microphone would be my first recommendation because the microphone element is closer to the source and further away from the table noise. A low sitting directional tabletop microphone (traditional/non-gooseneck) would be my second recommendation. If the ambient room noise is significantly high, I strongly recommend using a wearable type microphone. This will give you the best audio quality possible. The drawback of the wearable is that the end users have to be diligent enough to put the microphones on themselves for every conference.



As a tabletop microphone option, a gooseneck microphone would be my first recommendation.

If you can get the end users in the habit of using a wearable microphone and it's not a negative experience for them, they will have the best audio conference quality available today.

To Conclude

There are things that we can do as audio system designers to put ourselves in a position to succeed. Observing the space and considering reflection, absorption, and transmission is a good first step. Taking measurements of the ambient noise levels with a sound pressure level meter during peak hours can help narrow down the microphone options. Determine how strong the speech levels need to be captured in order to get good SNR. In my opinion, intelligibility is acceptable for conferencing when a room's ambient noise level is equal to or below 35 dB SPL and speech level is captured at a location where the acoustic



If you can get the end users in the habit of using a wearable microphone and it's not a negative experience for them, they will have the best audio conference quality available today.

energy is equal to or above 65 dB SPL. However, you should not take my word for it. You should experience the SNR ratios yourself and determine what you deem to be acceptable. A great technique that I recommend for audio conferencing designers is to create your own SNR reference bible. Gather noise level measurements in various spaces and create spoken word recordings within those spaces using various microphone types and placements. Listen back to the recordings while observing the noise and speech levels to determine what you consider acceptable SNR and good intelligibility. At what point is the intelligibility acceptable for you? If you can do this in many different spaces with various microphone types and placements, you will have created your own simple truths regarding acceptable speech and noise levels. Once you create that personal reference bible, you have legitimately done your homework and your end users will greatly benefit from having you as their audio designer. Attention to detail and intuition is what separates the great sound designers to the average ones. Good luck!

REFERENCES

Raichel - The Science and Applications of Acoustics (Daniel R. Raichel) 2000 Springer-Verlag New York, Inc.

Arons - A Review of The Cocktail Party Effect (Barry Arons) MIT Media Lab Cambridge, MA 02139

Figure 1 illustration created by Mark Waldrep and posted on www.realhd-audio.com in blog post titled, "More Detail at All Levels".

Figure 2 illustration found online at http://kf6hi.net/images/snr_example.jpg

Author's Note

I find it's best to simplify things when practical. The world today offers a massive amount of available information to us on any topic. We need to be able to categorize and de-clutter information to sort out the noise. This is what I have attempted to do with this topic of audio design for conferencing. My main goal is to offer the means to translate complex audio concepts into simple truths. As a designer, arriving at solutions by using your own simple truths is a sensible way to approach a project. Your confidence will be high which will instill trust in your clients and end users. Thanks for reading and good luck with your designs!



Ryan Shannon