A Theory On Loudspeaker Imaging by Jeff Bagby

http://techtalk.parts-express.com/showthread.php?t=195442

In 2001 I wrote a 3-part article that was posted on another website (now dead) with my thoughts on what factors influenced loudspeaker imaging. It resulted in a lot of discussion at the time. Due to the thread below I went back into my archives and found the article, and updated it a little with some current thoughts as well. I have posted the entire article below so it is very long. If you find it interesting please feel free to copy it off and keep it around. If you want to repost it on a website somewhere just make sure you keep it unchanged and give proper credit to it.

Thanks – Jeff B.

Part 1: Introduction and Theory

A while back someone wrote me and asked me to share my thoughts on what it takes to make a speaker that images well. Hmmm… that's a bigger question than either of us thought it was. Because after he asked me I started thinking about exactly how I would answer it and as a result my thoughts made a natural progression from one thing that influenced imaging to another. Suddenly I realized that several of these things were interrelated and that resulted in me arriving at a theory on loudspeaker imaging that I had not fully formed in my mind previously, nor is it one that I have heard espoused before (dangerous territory to be in, I assure you). The interesting thing about this theory is that it actually addresses a possible "scientific" reason why some people prefer analog source material over digital; why people who use speakers that use minimalist crossovers, minimum phase crossovers, and especially series crossovers say they image better; why Dipole speakers image well, why small mini-speakers image well, and even a few other things tossed in too.

Now, I am an engineer whose job focuses on problem solving, and I have to arrive at not only the root cause but also the best corrective action. As a result I am trained to look for patterns and relationships and to recognize that the probability that these relationships exist, or correlate, is unlikely to be by chance. This is the root of statistical problem solving. I guess I can't help but approach all questions that arise in this same manner to some extent. However, we need to keep in mind that although I use certain types of physics regularly in my profession, I am not a physicist, and I am certainly not an acoustician either, which means that my theory may be all wet. However, again, good problem solving techniques don't even care what field you are working in because statistical significance applies to nature in general. This is the basis of all statistics: DeMoivre's binomial and the resulting Fuzzy Central Limit theorem. So, as an automotive engineer applies his tools to speaker design here's my thoughts on what impacts a speaker's ability to image well. Someone, somewhere, may have researched this already (most likely has) and may have come to similar or very different conclusions, I don't know, but these are my thoughts.

The Theory:

First, we probably need to a good working definition of "imaging" and then we can build on that foundation and see where this takes us. I will define "imaging" as the loudspeaker's ability to recreate a sense of the original localization information that was present in the recording environment in such a way that vocals and instruments seem to be placed in space, even to the point of creating a sense of three dimensional depth. Will that definition work OK for everyone?

Now, let's take a look at what this definition implies. First, it implies that this localization information is present, which is not necessarily a given when you consider how many recordings are mixed. But when they are done correctly this information should be there. Second, it implies that there are aspects of loudspeaker design that influence, constructively or destructively, a speaker's ability to reproduce this information. My theory will only deal with implication number two and will assume that the criteria for implication number one has been fully met. I now present my theory:

In higher mathematics there is something called Fractal Analysis. A fractal is actually a mathematical expression that continues to produce a pattern into the infinitesimal. It is often demonstrated with electron micrographs of crystal edges. No matter how far you zoom in there exists a structural pattern along of the edge. This continues to the "infinitely" small, or in this case to the atomic level. I bring all of this up because I believe, well theorize, anyway, that localization cues within recordings are fractal in nature, because these cues consist of very low level secondary and tertiary, etc, delayed sound containing the acoustic "fingerprint" of the original environment. This information contains various levels of ambient frequency, phase, and delay encoding that reduce to a fractal expression. And this is precisely why certain conditions are either constructive or destructive in a speaker's ability to reproduce this information.

Unveiling lower level imaging information requires removing the multilayered veil introduced during sound reproduction that hangs over the sound in the first place. All localization cues, phase and path length relationships, and the "fingerprint" of the original acoustic space (as long as the recording is not mixed in such a way as to destroy this information) exist on the recording in progressive fractal levels and can be retrieved if these veils are stripped away. The key is to progressively strip away these veils, the effects of which are cumulative in nature.

For instance, analog recordings are by nature fractal in structure, but digital sources are not. This is because very low level information is still encoded within analog material, but with digital material there is an information "floor" and nothing at all exists below the floor. It would stand to reason then that analog recordings may have a higher potential of capturing and maintaining the low level localization cues than can possibly be maintained in a digital medium. How much can this effect things? I do not know, only that there is some unknown level of effect. There are still many people who swear that LP's played on high-end turntables create a sense of space that is missing from the digital world. And if what they are saying is true, then I theorize that the reason is due to the fractal nature of the analog source versus the non-fractal digital source. Possibly this difference can be heard and maybe because of this the digital vs. analog arguments carry some weight.

Here is a personal experience to describe this: Many years ago my wife and I were visiting a high-end audio salon. I was (foolishly) discussing the virtues of digital music and also my belief that expensive turntables had nothing to offer over less expensive ones. The salesman chuckled and asked if I would be willing to take part in a little experiment, to which we said, "sure". He proceeded to set up some music on a Rega turntable (which is actually a decent table) and we listened for a while. Then, through the same system, he played the same music on a Linn Sondek with an Ittok arm (I don't recall the cartridge, but believe it was a Linn as well). It only took a few seconds for both my wife and I to look at each other and exclaim the difference we heard. The sound was almost 3-dimensional over the Linn, it was virtually "flat" sounding over the Rega. I inquired as to how this could be, and the salesman explained that the suspension on the Linn made it possible to pick up much lower level information that contained this 3-D space. We then listened to a CD of the same music. It was even "flatter" than on the Rega. This is just one example. Localization cues are very low level on a recording, just as they are in life. If information retrieval is cut off before reaching this level a loss of imaging could result.

My theory is based on the fact that psychoacoustically we perceive the size of a sonic image by the relative intensity of the ratio of direct and reflected sound and the delay between the two. This is called the Haas Effect and it tells us that there exists specific thresholds for image broadening and ambience based upon this ratio of the reflected to direct sound. We also know that the direction we perceive sound to be coming from is based in part on the frequency response, or the transfer function of that sound. Therefore I theorize that the inverse is also true because it must hold to the same rules and there exist thresholds on the other end of the spectrum for the perception of localization cues as well. The problem is that the very things that tend to broaden an image also are the things that overshadow or veil the localization cues. I believe these are two different extremes on the same continuum.

I have already mentioned the difference between analog and digital source material, now in part two let me point out some veils that exist in loudspeaker design, how they affect imaging and how some designs attempt to remove that veil and restore the image cues (whether they realize it or not).

Part 2: The Room and the Cabinet

Essentially all of these culprits are things that "smear" the first arrival sound in such a way as to veil the subtle fractal elements of the original imaging cues. This refers to the Law of First Arrival Sound that says the first information to reach your ear will establish your perception of frequency response, direction, etc. In most of these cases we are dealing with external, additional sound sources. These additional sources are the veils.

One of the first ones that we all have to contend with are room reflections. We all know that reflections are secondary sources. In normal listening rooms they reach the listener with only a very small amount of delay. If this delay is less than 10 msec. we tend to interpret it as part of the direct sound. This is called the Precedence Effect, but that sound will be smeared due to the mixture of delayed sound and this will cause us to lose some perception of the subtle imaging elements within the recording. However, since all rooms have this problem to some extent, and most of us listen midfield in fairly dead living rooms, this item may not affect us as much as some of the other items do, but I am only speculating here. I know room reflections can cause big problems in response, so it stands to reason that the deader the room the better the speakers will image. This is probably where the LEDE (Live End Dead End) arrangement came from in the first place. There are a lot of devices out there designed to trap, absorb, and control room reflections. If you have a live room I would recommend trying them

Cabinet reflections and diffraction also affect first arrival sound due to their small additional reflected distances relative to the direct path length to the listener. Baffle diffraction is a major culprit in smearing the first arrival sound's localization cues.

I see two ways that loudspeaker designs attempt to address this problem. The most common way is to modify the baffle shape in an attempt to reduce the amount of diffracted and reflected acoustic energy. You see this in speakers like the Avalon designs with their faceted fronts, Waveform Acoustics with their egg enclosures, Thiel's curved design, Dunlavy's felt covered baffle, and many other brands that round edges or recess drivers. All of these designs attempt to reduce the difference between direct sound and diffracted or reflected sound. I also think it is generally accepted that speakers designed this way tend to image better than more standard designs. Just coincidence? My theory says maybe not, because if diffraction affects our perceived first arrival response then it has already contaminated that sound and affected the low level info as well. The higher level diffraction will cast a veil over the localization fractal information.

Another way this can be dealt with is by making a dipole speaker, where sound radiates equally forwards and backwards, but the front wave is 180° out of phase with the back wave. If you listen at the side of a dipole speaker, 90° from the front and rear axes, the sound cancels. Because there is almost no acoustic energy at the baffle edges of a dipole speaker to diffract, diffraction artifacts are almost entirely eliminated. For the same reason, dipoles also reduce room reflections from the sides too. Now, you may ask, "What about the rear reflected energy?" Well, that will arrive later than the first arrival sound. The effect it has on localization cues will depend on how much later. Therefore the closer the dipole speaker is to the wall the behind it, the earlier these reflections arrive, and the more they will smear low level information contained in the first arrival sound. If the dipole speaker is pulled out from the wall, delaying this reflected sound's arrival it may only add to the sense of space but not dramatically alter the localization cues. This relates back to the Law of First Arrival Sound. Therefore according to my theory, dipole speakers should image better, all other things being equal. Of course they seldom are, and dipoles have some significant problems of their own.

One of the areas in which dipoles have real problems, as well as all larger speakers, are cumulative frequency spectral decay cabinet resonances, which act similar to diffraction in that they are another external source of delayed sound. In this case we have a cabinet that acts like a big capacitor, storing acoustic energy, delaying it, and then releasing it in the form of a mechanical resonance. This resonance combines with the sound from the drivers, and like diffraction, smears the first arrival sound, veiling the subtle imaging cues.

Small speakers image better. Everyone seems to agree on this. However, the question is why. Most people assume it is because small speaker cabinets have less diffraction. But in reality the opposite is true. Small speakers have greater diffraction problems because frequency spread for the diffraction is usually narrower in bandwidth and the response issues are greater in magnitude. Narrower baffles do not have less diffraction than wider baffles, the only difference is that wider baffles result in diffraction at lower frequencies than narrow baffles. It can be argued that this is a good thing because our hearing is less sensitive in this lower range. So I believe that diffraction is not the reason why small speakers so often image better.

I believe small speakers image better because most small speaker cabinets are significantly stronger structurally than most large speaker cabinets. I think this likely makes the difference. Years ago I noticed that the speakers that kept getting the best reviews had one thing in common: They all had extremely well built cabinets that attempted to quiet as much panel resonance as possible. Consider the Wilson WATT with its lead-lined walls (in its original version, current versions use a special polymer material that is very inert to accomplish the same task), the 3" thick baffle on the Avalons, the cast marble baffle on the Theil CS5's, and the extensive bracing inside, etc. My theory says that by decreasing the cumulative cabinet resonance (less resonance magnitude, faster decay, less smearing) you will remove another audio veil hiding the subtle fractal information that establishes this sense of the original space that we are trying to recreate. The moral? Watch that structure and those mechanical sound sources, they are real and they will rob you of low-level information.

Part 3: The Crossover

Most loudspeakers are multi-way (at least 2-way) and as a result use more than one driver, with different acoustic centers, that produce different path lengths to the listener. These different path lengths are a source of time distortion. Sound from one driver (usually the tweeter) will arrive before the sound from the other driver (the woofer). Now, you may not be able to consciously hear this offset, but I believe it may be a source of another audio veil. Test me on this and try this experiment. Tilt your 2-way speaker back a little so that the path length for both drivers is very close to the same to your ear. Now, play some music and tell me if you hear more depth in the presentation. I bet most of you will say that there is. Every speaker has a Zero Delay Plane (ZDP) on some axis, bringing these voice coils into alignment on the listening axis will have a significant and measurable effect on step and impulse response. If the step and impulse are cleaner, then it stands to reason that the resolution of low level information might be better as well, there is simply less time smear present to affect it. The effect of aligning acoustic centers preserves arrival time cues across the whole spectrum.

The folks on the "Fullrange Forum" would swear that this is one of the reasons fullrange drivers are "superior" (in their estimation, anyway). However, having listened to a pair of Jordan JX92S's I can personally say that these things image better than anything else that I have had in my system. I cannot think of another reason for this other than the full spectrum is time and phase coherent. (Well, there is one more reason which I have yet to touch on, and that is the Jordan's lack of an electrical filter, which creates problems of its own).

Another area related but not identical to the time domain (path length) issue is the phase response of the speaker. Within their linear operating range, speakers are minimum phase devices, meaning their phase shift is directly related to the frequency response with no extra or "excess" phase included. There is usually a minimum deviation from the expected shift, hence the term minimum phase. But when you work with multi-way speakers the physical offsets and their resulting different path lengths, the acoustic roll-offs of the various drivers, their different phase responses, as well as the crossover used all alter this combined system phase response making it no longer minimum phase. For example, in a system with a 4th order Linkwitz-Riley crossover the frequency response may be linear through the crossover region but the phase response changes by 360°. This is excess phase. This phase shift is linear however, and will track between the two drivers, which we will address further down.

I believe the crossover design should preserve the phase response of the original signal as much as possible if our goal is to preserve the original signal as much as possible. In other words, all other things being equal, the more transient perfect the design, the better the sound. However, this is difficult and its necessity is even a bit controversial. There is a strong school of thought that says attempting to preserve the absolute phase response of the original signal is not audible and double-blind tests seem to indicate that they are correct.

I noticed years ago that some speakers produced a much deeper image with better localization than other speakers did. The first time was about 1980 listening to a pair of Dahlqusit DQ-10's (remember them). I have also noticed this effect from Theils, Vandersteens, Martin Logans, Magnepans, and Spica's just to name a few. Guess what these systems all have in common? Yep, they preserve the original phase response of the signal. I have listened to other very highly regarded speakers with excellent response that just did not produce this depth. They did not preserve the phase information. I think it does matter and it looks like there are still a lot of people out there that agree with me. I have heard a lot of people say, "All a minimum phase crossover does that's special is pass a square wave, and I listen to music, not square waves."

Well, doesn't it stand to reason that if it passes a square wave with low transient distortion then it would be for music too? Remember, we are after the retrieval of the low-level ambient and localization information. The ability to reproduce that square wave may be important here.

Now, referring back to the double-blind testing mentioned above, the problem with these tests, if they are performed with musical recordings rather than test tones, is that they still do not address one factor that makes all the difference: the recording itself. Most recordings consist of layers of sound recorded at different times, and even in different rooms, mixed together. If localization cues exist at all in these recordings it could be a confused mess of overlaid cues, which would be meaningless. But, do these tests with a proper Blumlien or Binaural two-microphone recording method and I bet people would begin to pick out the difference in depth of presentation between two speakers, one that preserved the minimum phase integrity of the signal and one that did not. This is just a thought. Testing could still prove me wrong, but some speakers that attempt to be transient perfect are known for the depth of their presentation and their imaging abilities. And if this transient perfect characteristic is a factor in this then what I have described, I believe, is the reason why.

But attaining this is a real problem for the DIY builder, so I will propose a compromise. You see it is nearly impossible for the DIY builder to construct a speaker that is actually minimum phase. Aligning the drivers and using 1st order crossovers will reduce the phase distortion but it won't get you home, and 1st order filters introduce other problems from over-driven tweeters or poorly suppressed high frequency break-ups in woofers, or just response anomalies in general. The problem lies in the speakers themselves. Take any good dome tweeter. They all begin to roll-off somewhere; most of the well-damped ones do this around 2 kHz. The roll-off could be 1st order from about 2 kHz to around 600 Hz where it makes a transition to 2nd order. Now if you add a single capacitor high pass filter to this you end up with a 2nd to 3rd order acoustic slope and an additional 90° of phase rotation that you do not want. Woofers usually roll-off rapidly above some point and this also introduces some phase shifts as well. In other words it is almost impossible to achieve minimum phase results with simple crossovers.

So the next best thing may be to preserve linear phase tracking between the two drivers, thus keeping them in relative phase with each other for a very wide range of frequencies. This is easily achieved by designing a 2nd or 4th order Linkwitz/Riley acoustic crossover. And I have found that I have been able to arrive at an acoustic 4th order L/R response with very simple electrical networks, generally using nothing more than damped 1st or 2nd order filters that arrive at flat frequency response and maintain great relative phase tracking. This being the case, I recommend the L/R crossover as a goal in designing for the DIY builder. Looking at a reverse null FR curve makes it relatively easy to verify that you have achieved good relative phase relationship between the two drivers. This is not to be confused with transient perfect because there is still some ringing, however in our quest for a better imaging loudspeaker we are incrementally reducing the distortions that veil our fractal information retrieval. Remember, these veils are cumulative. I will also note that many people feel the 2nd order version images better than the 4th order does. Could it be the reduced phase shift?

I have one more comment to make about crossover design and this one may explain why so many people say things like, "I switched to a series crossover and the depth and imaging just blew me away, it was night and day between this and the parallel network!"

Here's what I think may be going on here. Capacitors and inductors are simple reactive components, but they are also more than that because they all alter the sound in some way. If this was not the case we would all be using the el cheapo brand electrolytics and little iron core chokes and nobody would be spending $40 for an Audio Cap Theta if it didn't sound better (well some of you might, but not as many). The truth is these things do impact the sound. As a result my rule is to avoid overly complex electrical networks with lots of reactive components, keep the crossover as simple as possible for good results, and use high quality components when you do. From my own experience if you keep adding components to a circuit to correct for everything you can think of you will find that somewhere along the way your sound became compressed, flat, and lifeless. There is a cumulative effect to the degradation from all of those components. Let's use the K.I.S.S principle for crossover design (use your own words there, OK?).

This is why I believe people love 1st order circuits, especially series designs, even though the objectivist will tell you it shouldn't matter (the people who use them know better and I think their comments often support my theory). Series crossover circuits place these reactive components in different locations in the circuit and change what is in series with the drivers. Maybe this in turn reduces the thickness of the "crossover veil" and provides for better imaging cues to be reproduced. Don't get me wrong though, ALL crossover components are in the signal path even if they are shunt components, so all of them affect the sound in some manner (or they wouldn't be in the crossover in the first place.) However, their location and the amount of current they handle at their node may make a difference.

Remember again, these various veils are cumulative. Just like adding Saran wrap layers over a box, at first you can see right through and identify what's in the box. In fact, you would say the Saran wrap is clear. But add several more layers and you find that you can no longer make out what's inside. The cumulative effect of these layers has added up to veil our box's contents. It is the same way with all of these items I am listing. If we remove a few layers of the sonic saran wrap we may find our music begins to blossom into a beautiful 3-dimensional panorama of sound with instruments just hanging there in space. You may find that you can actually hear height and depth as well as width, and the width may extend past the speakers (I have heard all of these things).

I didn't even go into things like the M.A.R.S. system that Irving M. Fried used in his last incarnations and the old Carver Sonic Holography and the Polk S.D.A.s. All of these operate on the same principle of canceling interaural crosstalk between the two stereo channels and your two ears. I have designed a system that did this very simply and very effectively and on some recordings the effect was almost unbelievable, talk about 3-dimensional, wow! But we can save that for another day. For now let's just focus on reducing the veils that cover our ability to retrieve and reproduce the complex yet subtle fractal ambient and localization information that exists in the recording.

Additional thoughts:

One of the things not covered well in the original series was the fact that a speaker should not only have a flat frequency response in first arrival sound, but it would be best to have smooth power response free of series dips and peaks as well. In addition to this; for proper stereo imaging it is imperative that the response of the Left and Right channels be as much the same as possible. Deviations in response that differ between the two sides can shift the image from one channel to the other in certain frequency ranges. Because of this, speakers that are carefully designed so that the left and right speakers are as identical as possible will be candidates for good imaging. This should have been covered more in the original series.

Something that is a hotbed of discussion these days is harmonic distortion in drivers. It would stand to reason that if resolution of fractal information is a key to recreating the original space of the recording then lower distortion would be a benefit in retrieving low level information. At this time, I do not know how much this matters. I know I have heard good imaging speakers that did not use the finest drivers, so I have to consider this too.

Another thing I didn't discuss in the original article above was attempt to rank the relative magnitude that each of these potential contributors hold. So what you think they are? These are just my personal thoughts, and I have never done any testing to confirm this one way or the other, but here's what I think based on systems I have listened to:

1. First and foremost, I think having the correct tonal balance and near perfect balance between the Left and Right speakers have to be near the top because frequency response dominates our perceptions of speaker performance. And how well diffraction is controlled is also a factor in the smoothness of frequency response so it will be a part of this one too.
2. Second, I believe the mechanical inertness in the enclosure comes next based on the systems I have heard and the characteristics that they shared.
3. And third I think are the time domain issues. This is less important than the first two issues because its audibility is lower. However, if everything is done right I still believe that a transient perfect speaker will offer advantages over the speaker that has a lot of error in the time domain.

Jeff Bagby
Original 6/14/2001
Revised 3/10/2007