Testing One, Two, Three

Stuart Yaniger

Have you ever watched a talented close-up magician? Coins seemingly appear and vanish in defiance of the laws of physics. Handkerchiefs are cut to bits, then are made whole. Bullets are caught in teeth. Amazing! And watching the show, it all looks real, even when your conscious brain KNOWS that the magician is trying to fool you- and if the magician is skilled, even when you know the trick, you're still fooled (I used to work with a fellow whose Three Card Monte would buffalo experienced street hustlers). Our brains and sense organs, as a system, are limited in what they can do and how they can efficiently process the sort of data that kept our ancestors from being eaten or trampled. Magicians know how to exploit the sensory system's "shortcuts" and use that knowledge to entertain.

Unfortunately, that knowledge can be used by charlatans and fraudsters- if the fellow bending spoons or the lady talking to the dead tell their audiences that it's not an act, it's real, they can be quite convincing to the majority of people who do not have experience with the arts of conjuring or cold reading. As with the ethical performer, the audience can be fooled into believing that they saw or heard things that they didn't actually see or hear. Does that make the audience stupid? Not at all. They're human, with all the basic hardware issues that go with having a human brain.

The difference between the audiences of a magician and a charlatan is not intelligence or perceptiveness but rather gullibility and ego. The audience at a magic show knows they're being fooled and will not walk away convinced that the magician really materialized doves out of thin air or read their thoughts. The audience for spoonbenders and channelers often think that they are too smart to be fooled. Unfortunately, being fooled by one's senses is universal, and many scientists have had careers ruined by endorsing transparent frauds (1). One magician I know told me that he LOVED having scientists and academics for an audience- "They're much easier to fool than children or bus drivers since they rely on their intellect and always assume they're the smartest person in the room. And they usually are, but they're not working in their environment, they're working in mine."
It's important, when trying to get to objective truth about subjective judgments, that one's ego is left standing outside. This isn't about who's smarter or whose ears are better, it's about determining what's important in your designs and making design decisions with ones, errr, eyes open. And that requires controls on our listening tests.

1. Our Purpose

"I changed the cathode resistor in the first stage and the difference was staggering!"
As the wise philosopher Alvarado asked, "What is reality?" (2) Philosophers have been philosophizing over this question for millennia, yet have come to no definitive answer. Thus, we can safely and swiftly leave them to their own devices. With the philosophers safely receding in the rear-view mirror, we can now ask the fundamental question of concern to the audio designer, "Is doing X instead of Y really an improvement? Can I actually even hear a difference between X and Y?" The naïve and obvious answer is, "Trust your ears. If you hear it, it's real." And like most obvious statements, it's a truism. And that's the object of this article- how to evaluate sonic differences at a designer level with accuracy and confidence, to KNOW what's coming from the sound and what's coming from other factors. The important thing in audio is what you HEAR, not what you hope you hear, believe you hear, or think you hear.

The rational audiophile will accept as axiomatic the proposition that if the 3 dimensional soundfield of an acoustic event is reproduced accurately, the listener will be unable to distinguish it from the original- his ears and brain will be fooled. In reality, this is impossible with current technology (where the 3d soundfield around the listener is collapsed to two to five points remote from the listener, each having a single valued function of time), and audio systems must rely on tricks and illusions to fool the listener. The best-known "trick" is stereo- the center image is a pure illusion, there really is no singer standing between the two speakers!

If, like the magic show, the goal of a hifi audio system is to provide entertainment by creating an illusion, there's more than just the pure soundfield to be considered. But the informed designer needs to know what counts for the actual auditory parts of that sonic illusion and separate that from the rest. That's the purpose of this little exposition- controlled subjective testing is a powerful tool which has contributed hugely to audio in general. It is a method and philosophy that can be inexpensively and (relatively) easily used by individual designers and experimenters.

2. Define the Problem

When we are trying to evaluate the subjective sonic impact of a design choice, we need to first define what it is we are trying to find out:
1. Can the consequences of this choice be heard by me?
2. Can the consequences of this choice be heard by an average listener?
3. Can the consequences of this choice be heard by *anyone*?
4. If the answer to 1, 2, or 3 is "yes," then which choice is preferred?

This seems trivial, but it's not- it's a natural tendency among audiophiles to want universal answers. As a result, the boundaries between these questions get blurred- each of these questions is quite different from the others, and naturally, the means by which they can be answered is also quite different. In particular, Question 4 is a distinctly different one when applied to the population at large

 Free download from www.linearaudio.net © Linear Audio and Stuart Yaniger

or the target market- the usual term for experimentally determined preference is "Hedonic." Getting reliable hedonic data is a massive undertaking- an excellent example in the audio realm is the work by Toole (3) in determining listener preference for loudspeaker frequency response and polar pattern. In this article, we'll look at methods appropriate to answer the question appropriate to the main issues facing high end audio designers- Question 1. We will touch on Questions 2 and 3, but only long enough to determine that this is the sort of issue best left to others- our focus will be in properly framing the questions and designing experiments appropriate to answering those questions- and avoiding the error of designing a test to determine the answer to one of the questions, but really wanting the answer to a different one (this is far more common than one would expect!). I'll also review some common ways in which typical audiophile approaches to listening tests can be misleading and how to do it better.

This is *not* a guide to testing the claims of others (except in a limited sense), to do scientifically rigorous sensory analysis, or to determine the limits of human sensory performance, but we will borrow from the methods that researchers use in these areas (4) to allow you to get useful results that will allow you to focus your design efforts on things that matter- that is, matter to you, the designer or to the people that you design for. The false dichotomy of "subjectivists" and "objectivists" is one of the most destructive notions that high quality audio has ever had inflicted on it. That notion has unnecessarily poisoned useful discussion of design issues and led to a comic book view of rational approaches to making real design improvements. Subjective auditory impressions, when controlled to be by ear alone, are hardcore objective data.

The assumption here is that, as a designer, you want to get the right answer, not the answer you're hoping for (unless it's right). If you are testing strangers or groups or there is some interest at stake by the listeners, controls have to be structured quite differently, beyond the scope of this short article.

We start with a quick review of what we're trying to accomplish, the ways we unwittingly sabotage our determinations of audibility and preferences, talk about how listening tests should be structured, then give a few case studies as examples. The assumption is that you don't want to run full University trials or the complex methods outlined in the standard BS.1116-1 ("Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems"), that you have a couple of friends- at most- to help out, and that you either want to satisfy your own curiosity or drive design decisions.

Remember: Trust your ears, but don't trust your lying brain!

3. Non-Auditory Cues

As we already touched on, what we are interested in finding out is if a design choice is audible. Whether it looks cool, comports with audio fashion, satisfies philosophical urges, gives you pride of ownership, or it just makes you happy to have when playing your music are separate design issues beyond the scope of this short article. But fundamentally, to evaluate audibility, we have to trust our ears, and not help them out by using other senses or our conscious or unconscious preconceptions. We will begin by reviewing some of the principal ways we cheat ourselves into hearing things that may not be there and how to avoid them.

3a. Placebo Effect

I don't think we need to rehash the placebo effect- just about everyone is familiar with it. You perceive a change because you *think* there's a change, whether or not there's been one. It's programmed into how our brains work, and is probably a result of the relative impact of false negatives (I thought I heard something, I figured that I was imagining things, did nothing, and got eaten by a sabertooth tiger) and false positives (I thought I heard a sabertooth tiger, there wasn't one, so I ran for no reason) on our distant ancestors. This is the major reason why any serious auditory (or haptic or organoleptic, for that matter) inquiry must be done blind, i.e., that the test subjects- whether you or your Trusty Assistant- are unaware about whether or not there's been a change of the variable under test except by their subjective reactions to the sound. For example, we love big, heavy, shiny, artfully designed boxes for our electronics. But does the cheap, flimsy amplifier actually sound different than the 100 kilo three chassis monster? Maybe, but if we see them both, the answer will almost assuredly be "yes." So to determine whether the actual sound is different, it's necessary to hide their identities during audition. You may know that the comparison is between a $100 Flanasonic DX-5000 and a gorgeously constructed $10,000 Dominator Beethoven Mark IV, but at any given time, you don't know which of the two units you're listening to. If they sound different, your ears will tell you. If they don't, your ears will tell you that as well.

One interesting corollary to this is the observation by several cynics that amplifiers with black front panels tended to have their sound described as "dark" when compared with amplifiers having light colored front panels…

Listen with your ears, not your eyes.

3b. My Wife, or The Need for Double Blind

It's become an audio cliché- "The change was so obvious that my wife, who wasn't even in the room and doesn't care about audio, immediately asked me, 'What did you change? It sounds better.'" And that does happen- as I will relate shortly, it's happened to me. The cliché provides an insight into a unique way in which you fool yourself with the unconscious help of others.

Many years ago, a German teacher named Wilhelm von Osten had a horse, Hans. Hans was particularly notable because, it was claimed, he could do arithmetic. If he was asked, "Hans, what is four plus three?" Hans would respond by pawing the ground seven times. He could do all sorts of arithmetic, not just addition. That indeed is remarkable, for it demonstrated not only mathematical ability, but also language capacity and understanding. By all accounts, von Osten was an honest (though ill-

 Free download from www.linearaudio.net © Linear Audio and Stuart Yaniger

tempered) man, so no guile was suspected. Hans's performances attracted large crowds, and the high public interest prompted a psychologist, Oskar Pfungst, to investigate.
Pfungst noticed several things. First, the trick worked even when von Osten wasn't present- this eliminated the idea that von Osten was consciously signaling the horse. Second, the horse wouldn't perform properly when it had blinders on. Third, and most tellingly, the horse wouldn't be able to answer questions that the questioner couldn't.

Now, how does this relate to the wife in the next room? I had that experience myself, several times. When I would fiddle with some components or circuit parameters, and sat listening joyously to the results of my creative labors, my wife would often notice as well that my ministrations had been effective. Clearly, with this unbiased ear (she knew nothing whatever of electronics or acoustics) validating my impressions, I must be on the right track… or so I thought, until one evening, some doubts crept in. Thinking uncomfortably about a variable that I hadn't controlled, I changed nothing in my sound system, but managed to get a, "Did you change something? It sounds nicer," from the next room. A pin to my ballooned ego. How did I prompt this comment without changing anything in my system?

Simple. I played a cut from "Jazz at the Pawnshop" and a cut from "Kor," both very popular "audiophile" albums of that era, brilliantly recorded, but of somewhat limited musical interest. I had fallen into the habit of using these for evaluating my system, but wouldn't ever play these for the sheer joy of hearing boring music. Completely unconsciously, I was cuing my wife that Something Changed and that I was Seriously Evaluating. And being that she liked to say things that made me happy because of the positive feedback to her, her response to those two cuts was to say something nice about the sound. Nothing conscious on either of our parts, mind you, just a lovely *folie a deux* that happens far more often than many of us would be comfortable to admit.
The point of this is that, even when you've been careful to hide the obvious variables, there are ways of cueing others, even when they can't explicitly see what's being changed (5). Beware the unintended variable!

For this reason, a valid test will be double blind, that is, any persons in the presence of the listener need to also be unaware of what among the choices is being heard. The experimenter can easily cue the listener with neither of them being consciously aware. If components are manually switched in and out of a system, it has to be either done out of sight of the listener, or the listener has to exit the room while it's being done. Naturally, the person doing the switching will need to leave the room before the listener returns.

3c. Timing
When equipment is interchanged, time is of the essence. This has two meanings- first, the human auditory memory is rather short. If it takes 20 minutes of fiddling to change from one component to another, your memory of the first component will be quite imperfect. Your mood may subtly change,

the barometric pressure may change… it's important to try to keep switching time as fast as possible for any direct comparison testing.

The second meaning is a bit more subtle. Say I'm comparing component A to component B. For each trial, my assistant randomizes what I'm listening to. If it's truly random, half the time, the next presentation will be the same as the last, i.e., A will change to to A, half the time the next presentation will be different, i.e., A will change to B. If the assistant takes 10 seconds to determine that there's no change needed, but 2 minutes to change something, I can tell whether or not there's been a change from the time lag. This is a bit of an extreme case, but in order to prevent more subtle non-auditory cueing, the assistant should actually disconnect and re-connect A (or B) even when the next presentation is unchanged from the last. Besides eliminating the timing cue, the other salutary side effect of this procedure is to ensure that any switch contacts or jacks/plugs are continually exercised.

3d. Level

There are lots of ways to consciously or unconsciously bias a listening test. I've run across several of them… A favorite trick in hifi showrooms is level. Curiously, small differences in level (<1dB) are not generally perceived as such. Our ear/brain tells us that the slightly louder choice sounds "clearer," "more open," or a similar descriptor. It doesn't take much for sensitive listeners- some people can detect level changes as low as 0.1dB. So the salesman confidently offers the potential customer a switchbox and tells him to compare the cheap hifi component with the more expensive one (which the salesman has judged that this mark can afford). The level on the more expensive unit is just a wee bit higher. Aided by the salesman's encouragement, the more expensive unit DOES sound better and the checkbook is opened.

As designers and experimenters, we don't want that simple variable to interfere with what we're looking for. So it's best to match levels to the limits of what expert listeners can hear lest you fall for your own inadvertent sales pitch. It's fine to make comparisons where you can change the volume, but the volume of both components being compared for sound needs to be varied simultaneously and equally- no matter what the system volume, the volume of the two components be matched.
The same goes for frequency response- often, this is an effect of component changes and should be measured before doing any listening tests. If the test is to determine audibility of something other than a frequency response or level change, these should be equalized before beginning the controlled subjective trials.

3e. Miscellaneous

A particularly amusing cue, which actually WAS sonic (though not as intended) was related by Professor Stanley Lipshitz (6). In this experiment, a first generation digital processor (Sony F1) was inserted into an audio circuit and compared to a bypass using a relay-driven switching box (7). The test listener, who was a prominent critic of digital audio (and just happened to be a manufacturer of turntables) couldn't hear the difference with the Sony in or out of the circuit. Lipshitz could, detect-

 © Linear Audio and Stuart Yaniger

ing a small change in the background noise- and that's absolutely a valid and audible difference. Lipshitz's colleague noticed one other thing, which in a sense was an auditory clue- the relays that switched between the Sony and the bypass were located in different positions on their chassis and thus made a slightly different sounding "click" when they opened and closed. He could then tell which unit was switched in when there was no music playing by switching back and forth and listening to the different tones of the clicks. Oops.

4. Test Formats

It is important to keep terminology straight- many audiophiles casually interchange terms like "double blind test," "objective evaluation," and "ABX" despite these terms having very different meaning. As we discussed above, any valid test of audibility (a subjective phenomenon) needs to be controlled for non-auditory cueing, and that generally means double blind. Double blind listening tests are both subjective and objective- if you can subjectively distinguish A from B ten times in a row, you have objectively demonstrated that they subjectively sound different (the converse is not true- can you see why?). Double blind tests can be done in many different formats, not just the well-known ABX- we will outline a few of them.

This would be an appropriate place to talk about statistics, but I won't. There are some excellent treatments of the basic statistics needed for test design in about a million textbooks. A particularly good online treatment applied to listening tests is given in Reference (4). In fact, the entire site is well worth reading. In a nutshell, a single positive result means very little, usually no more than a single coin flip. Two positive results means more, two hundred mean a LOT more. The choice of the number of repeated trials is a tradeoff between fatigue factor of the listener and the desired certainty. In most sensory work, the number of trials and required number of correct answers is chosen to give a 95% confidence that the results are not due to pure chance. For example, if the chances of a correct answer purely by chance are 50% for each trial (like coin flips), a set of 10 trials will require 8 correct to have a 95% chance of being other than random (actually, it's slightly less than 95%; 9 out of 10 correct means that it's 99% likely that the results are not due to chance). An excellent set of calculators for a variety of test setups may be found at www.stattrek.com.

One side note: statistics of large groups mean that if you have 100 people making random guesses, the chances are good that at least one or two people will get 9 or 10 out of 10, even if there is no audible difference. Similarly, if listeners are biased to want to hear differences (very common among audiophiles!), it is very important that the test statistics account for that fact if a test is structured for people to guess "same" or "different," they will prefer "different," even absent any differences. This is not a problem for other test configurations.

4a. ABX

In the ABX test format, the two different devices to be compared are labeled as A and B and connected to some sort of switch box. The listener is free to switch between A and B to get a handle on any sonic differences. The third switch position is X- randomly chosen to be either A or B. The job of the listener is to determine whether X is A or B. The test can be structured in many ways- the listener should always have switching control, and can have control of volume, source material. There is no inherent reason why the listening and judgment need to be rapid- with proper controls in place, they can extend over days or weeks, if you desire.

There have been commercial hardware implementations of the ABX tests (7), which have been used to demonstrate high listener sensitivity to minor changes in level, frequency response, absolute polarity, and phase shift (6, 7). Although some have objected to the use of this particular hardware implementation, there has been no demonstration that their inability to distinguish particular electronics is due to a flaw in the Clark hardware rather than the possibility that they indeed cannot distinguish between the electronic devices under test. Nonetheless, if a particular hardware implementation of the ABX test is found wanting, it is not difficult to construct an ABX box with whatever switches or relays are deemed adequate for audiophile use. After all, the signal in an audio system has passed through many switch and/or relay contacts during recording and mastering, and most audio systems have source selector switches. More on this below.

An important point to note is that in each trial, the chances of correct identification even if the two choices are indistinguishable is 50%. Hence, multiple trials are needed to achieve any sort of significance. The more trials, the better for the statistics, but worse for the listener- once the interesting aspect of the listening test has passed, the test can easily become drudgery, and subjects get easily bored.

4b. Triangle

The triangle test is very commonly used in organoleptic (smell and taste) analysis and is nearly universal in wine evaluation. In a triangle test, the listener is presented with three choices, A, B, and C. Much like the old Sesame Street song, "One of These Things is Not Like the Others," two of the choices are the same, one is different. The listener must choose which of the choices is the odd man out. One advantage to this method is that it is statistically more powerful than ABX, that is, it takes fewer trials to establish significance. It should be used more in audio, but it isn't.

4c. Paired Forced Choice

The Forced Choice is a powerful and effective tool for determining thresholds of subjective perceptions by basing the next presentation on the outcome of the last one rather than randomizing. It typically requires a low number of trials to achieve significance, but does require some dedicated assistance- keeping in mind, of course, the need for double-blind (unless your name is Hans). It's best to use for situations where the test component is inexpensive or can be easily duplicated. It's the first choice if the sensory variable under test is amenable to software simulation (e.g., level, frequency response, distortion, compression). An example to illustrate the use of Paired Forced Choice is given below.

 Free download from www.linearaudio.net © Linear Audio and Stuart Yaniger

4d. Sorting

This is most useful for phenomena which can be captured on digital files or can be easily replicated. There is a population comprising two different sorts, A and B. The listener needs to separate the population into two piles. The number of A and B in any population needn't be equal. We show an example of how this is done below.

5. Hardware Solutions

5a. Case Study: Coupling Capacitors

The dielectric materials and construction of coupling caps gives audiophiles nightmares- or more cynically, a way of entertaining themselves for years. By and large, measurements have shown that one can measure differences between capacitor types in very particular applications, for example, high current or DC timing (8). How well do these measured differences translate into audible performance of coupling caps, where DC accuracy is not an issue and currents are low? Do hyper-expensive "audiophile" caps actually improve the sound? Can you even hear the difference between a super capacitor and a cheap piece of junk in that position?

This was a question on my mind and led to the construction of the Bastard Box, the schematic for which is shown in **Figure 1**. Basically, the Box is a method of doing a double blind comparison be-

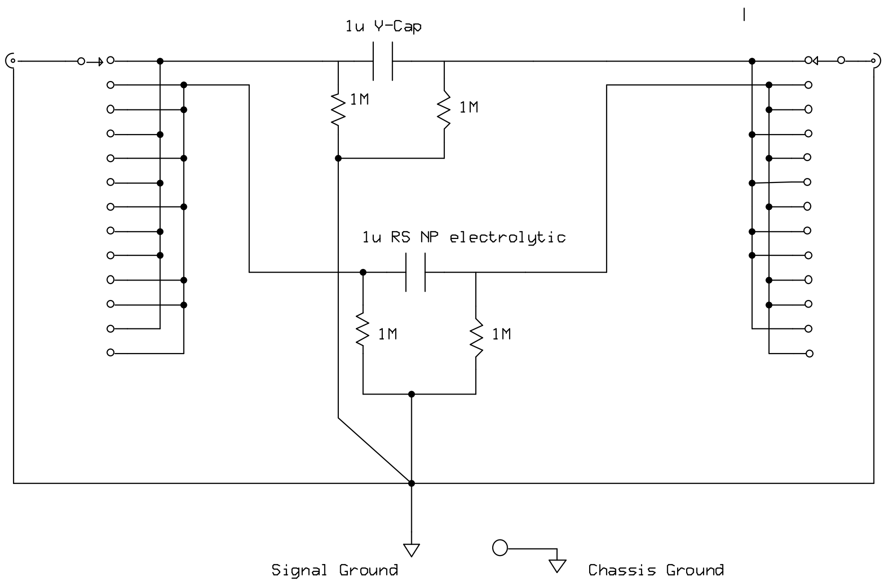

Figure 1. The Bastard Box. In this configuration, Position 1 is a V-Cap Teflon, Position 2 is a Radio Shack electrolytic, and positions 3 through 14 are randomized.

tween two types of capacitors without worrying about level matching, allowing the listener to choose source material, switching times, and can be easily inserted into nearly any sound system at line level. The Box comprises a pair of high pass filters with the same cutoff frequency allowing different capacitor types to be auditioned with a constant frequency response and level. To make the comparison as easy as possible, the two capacitor types being compared were a very high priced Teflon dielectric capacitor (V-Cap TFTF, about $300 each) and a cheap bipolar electrolytic (Radio Shack, about $1). The caps were confirmed to be within 5% of nominal value- matching COULD be tighter, but it's doubtful that the minor changes in frequency response at subsonic frequencies will be audible. The values given here are chosen to be compatible with my power amplifier's input impedance, 100k. With the values shown here, the 3dB down frequency is roughly 2 Hz, low enough to prevent excessive phase shift at the low frequency limits of my system (20 Hz).

The box is fitted with premium RCA plugs for input and output, is wired with audiophile-approved silver wire insulated with Teflon, and internal connections soldered using silver-bearing solder. The switch is a massive Cinema Engineering 14 position three pole rotary switch, with silver contacts and a leaf-spring washer. Overkill for sure, but you don't want to have poor quality switch contacts obscure the differences we want to listen for. The resistors diminish the tendency for clicks and pops to occur when switching and perhaps inadvertently cue the listener.

In this circuit, position 1 (A) corresponds to the V-Cap in series with the signal. Position 2 (B) corresponds to the cheap electrolytic in series with the signal. The next 12 positions of the switch (X) are randomized- I used coin flips to determine which of the two capacitors would be switched in-circuit for each switch position, then wired accordingly. The schematic as shown can be (and should be) modified to reflect whatever random sequence you come up with. Coin flips work well for this; an electronic version may be accessed at http://random.org/coins. Another way to get a random sequence for A and B of length n is to go to a table of the decimal places of pi (or some other transcendental number), pick a spot at random, then for the next n numbers, assign even to A and odd to B. Note that doing a randomization by arbitrarily picking "A" or "B" is not as good- humans feel that having too many As or Bs together spoils the randomness. Yet Nature is just as likely to pick an A following an A as it is to pick a B following an A.

If you're trying to test your own perception, have a Trusty Assistant connect the randomized positions in your absence and keep a key sheet. You now have all the time you like to try to determine the identity of each switch position. Record the guesses, then hand your score sheet to your Trusty Assistant. A little bit of statistical crunching and you'll soon know if you really could hear a difference. The test box can just as easily be configured for a triangle test by having positions 3, 4, and 5 be one triangle, 5, 6, and 7 be another, and so on. The box can also be used in a sorting mode where the listener groups the sonic impressions of the different switch positions.
Can the presence of the switchbox (switch contacts, extra interconnect, plugs and jacks) obscure the differences one is trying to hear? To determine this, before installing the capacitors, I wired several

 Free download from www.linearaudio.net © Linear Audio and Stuart Yaniger

of the positions directly with a short length of silver wire. This put all of the "extras" into the signal path, other than the capacitors. I manually connected and disconnected the box between my preamp and power amp. There was no difference that I could hear. If you do the same and think that connecting the switchbox is degrading the sound, this can be easily confirmed with the help of the Trusty Assistant. Arrange some sort of visual barrier so that you cannot see whether or not the switchbox is present. Leave the room. Have Trusty Assistant flip a coin- if it's heads, he should connect the box, if it's tails, he should bypass it. After a predetermined time (4 or 5 minutes is usually more than adequate), he should exit the room and you can return. Can you hear the degradation? Mark a scoresheet, then repeat the process several times. If indeed you can identify the presence of the box, you can try to determine where the malfunction is and correct it.

I spent a week playing with this setup comparing capacitors and marked my scoresheets. I scored 7 out of 12 correct identifications. I'll be interested to see if others can do better.

One interesting variation of this test involves only changing the variable in one channel of a stereo pair, leaving it fixed in the other. Our ear/brain is highly sensitive to localization (the direction the sabertooth tiger is coming from is important to know in order to prevent being eaten)- if two channels in a stereo setup are not perfectly symmetrical, the image or soundstage can be smeared from its intended position- in some cases, an instrument's location can seem to wander depending on the note being played. To look at this effect, the Bastard Box was wired to have just the V-Cap in one channel and switch between the V-Cap and the electrolytic in the other. My score was 8 out of 12- a little better than before, not quite at the point of significance, but close enough that I want to try this again. The use of switching presentations in one channel only to take advantage of localization abilities is another ripe area for designers and experimenters to play with.

5b. Paired Forced Choice Case Study: The Audibility of Op Amp Buffers

This case study illuminates an interesting technique in sensory analysis, the Paired Forced Choice.

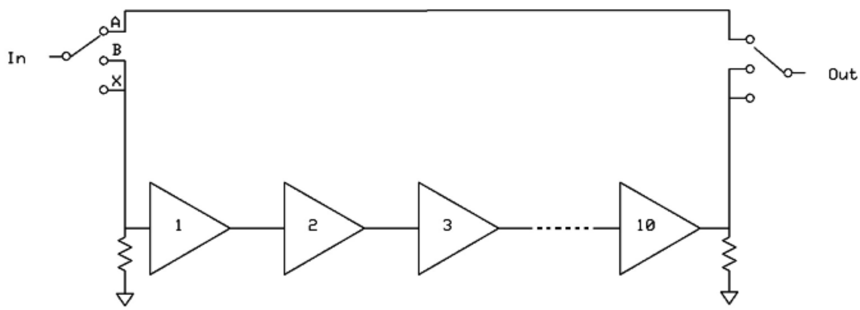

Figure 2. Buffer Test Circuit. Position A is a straight wire, Position B is a chain of up to ten buffers, and Position X is randomized. In the shown configuration, Position X selects the chain of buffers.

I was interested in using an IC op amp in unity gain mode as a buffer in a mike preamp that I was building. I admit to a bit of a bias against them- they're seemingly too perfect and definitely too easy. Would it foul the sound to my ears? I enlisted a Trusty Assistant and proceeded to find out. I put together a simple prototype circuit, then built nine more! Placing all ten in series between my preamp and power amp, I fed music into the input and listened to the output compared to a wire bypass around the chain of buffers - I was certain there was a difference! To confirm that I did, we wired up a three position switch as shown in **Figure 2**. As before, A was the chain of buffers, B was the straight wire bypass, and X was one of the two (I did not know which in advance). It was easy to identify X by ear.

Trusty Assistant then removed one buffer from the chain, leaving nine connected in series, then re-randomized X by a coin flip. Again, no problem identifying which was which by ear. Trusty Assistant then removed one more buffer, leaving eight. Harder this time, but I got the right answer. Trusty Assistant, not surprisingly, removed one more buffer, leaving seven. This time, I guessed wrong. Trusty Assistant added the last buffer back, bringing it back to eight. I guessed right. Back to seven, I guessed right, so we went back to six. I guessed right. Down to five. Wrong. Back to six, right. Down to five, right, down to four, wrong. Back to five wrong. Back to six, right. A diagram of these trials is shown in **Figure 3**.

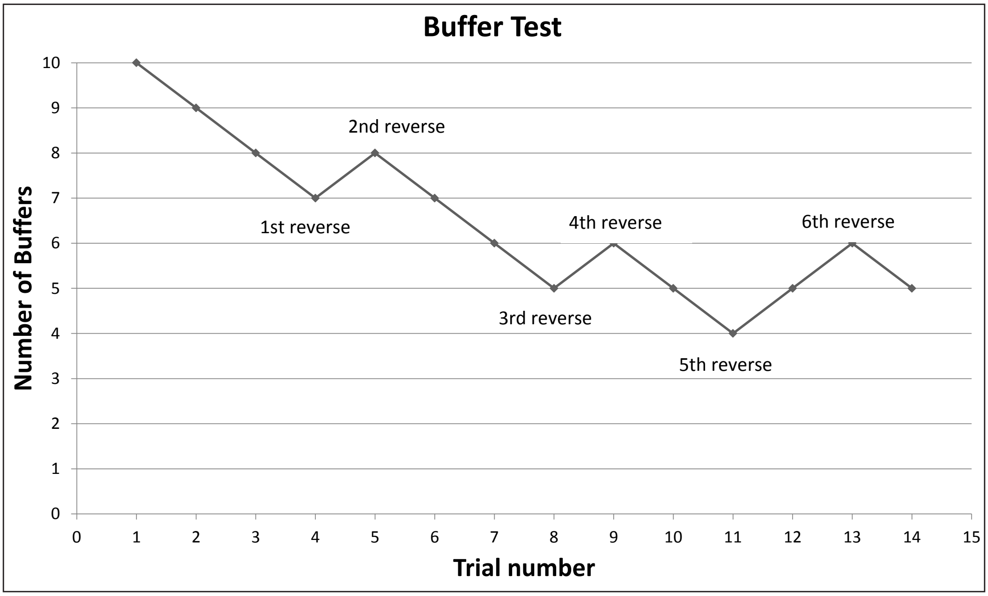

Figure 3. Plot of Buffer Test results using the Paired Forced Choice method. Note that the detection limit appears to be somewhere around 5 buffers in the chain.

 Free download from www.linearaudio.net © Linear Audio and Stuart Yaniger

I think you can see the pattern- every time I guessed right, the number of buffers was reduced, when I guessed wrong, the number was increased. By prior agreement, we decided to end the test when the direction of the addition/removal of the buffers changed six times or until we ran out of buffers to remove. So if, for example, I could reliably detect a single buffer, that would end the test. If I were able to reliably identify the presence of six buffers, but couldn't reliably identify five, my answers around five buffers would be random and thus frequently switch directions (analogous to a random walk!).

In this case, my threshold seemed to be somewhere around six buffers. So I felt pretty safe using just one in the signal path. As a reward for the first reader to guess what I heard and how I identified the buffers, I'm offering a free, autographed portrait of Joe M. Varilla.
This shows the usefulness of Forced Choice in getting information beyond "can/can't hear" with a relatively small number of tests.

5c. Listening Time Case Study: Preamplifiers

One commonly made assertion is that Technology X (fill in the blank with your favorite) reduces listener fatigue. The audiophile spends more time listening and is more relaxed. This is not something that will be captured in a relatively short conventional blind test. I had two line stages on hand which seemed like they would be good candidates for such a test- their measurements were excellent and in a short A/B sighted comparison, I couldn't tell them apart. I also had some surplus hour meters, similar to what one finds on industrial equipment to let the operator know how many hours the equipment had been in service (useful for periodic maintenance). At the start of every listening session, an electronic coin flipper (bought as a science fair kit) activated a relay which switched the signal through one or the other of the line stages- the volume control was common to both and under my control. Each linestage got its own hour meter to measure how long it was used.
I left this set up for about two months, then read the meters. One had about 15% greater elapsed time than the other. Is that significant? I don't think so, but this is a rich area for experimentation (several approaches to a hedonic measurement come to mind), and I would love to hear about your results if you try something like this.

5d. Sorting Case Study: Wire Directionality

This was quite a fun one, but with no resolution so far. Nonetheless, the test setup is instructive. A question arose about the audibility of the direction in which a standard coaxial interconnect was connected. Conventional engineering would have it that for audio signals, the interconnect "direction" should have no effect. Yet reports persist that this might not be the case. One person in particular was insistent that reversals of interconnects were clearly and ambiguously audible.
A reel of coaxial wire was obtained and ten interconnects were assembled using standard RCA connectors. The interconnects were labeled individually and the direction of the cable with respect to the winding direction of the reel was noted for each interconnect. One end of each interconnect was randomly marked and a key sheet was kept as to whether the mark was on the "leading" or "trailing"

end. The interconnects were then sent to the listener to sort. Unfortunately, the listener decided not to proceed with the test, so this one is still ripe. I can say that I can't hear any difference with wire or cable direction, but if someone can sort them in a test like this, it may well be an issue worth considering.

6. Software Solutions

If the presentations are amenable to digitization (for example, the effects of EQ or compression), then .wav files can be conveniently compared in an ABX format using the foobar2000 media player (available for download at www.foobar2000.org). This is an amazingly convenient way to do double blind comparisons!

7. Wrap-Up

If you think you hear a difference between two things and want to KNOW that you or your audio buddies really are hearing something, and then use that knowledge to optimize the sonic aspect of your designs, controlled subjective testing lets you isolate the sonics. The oft-heard canards ("There's test pressure!" "The switchboxes are no good!" "The switching is too rapid!" "The reference systems suck!") which are used to object to trusting ones ears are easily accounted for- IF the designer or experimenter wants to. Doing a controlled subjective test is relatively simple, inexpensive, and usually needs no more than one Trusty Assistant. With only a moderate effort and expenditure, a designer can quickly get good actionable data.

References:

(1) Randi, James. 1982. "Flim-Flam" Prometheus Books.

(2) Proctor, Philip; Austin, Phil; Bergman, Peter and Ossman, Dave. 1970. "Don't Crush That Dwarf, Hand Me the Pliers," Columbia C 30102.

(3) Toole, Floyd. 2008. "Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms," Elsevier.

(4) Martin, Geoff. 2011. "Introduction to Sound Recording," available through www.tonmeister.ca; Lipshitz, Stanley and Vanderkooy, J. 1981. "The Great Debate: Subjective Evaluation", J. Audio Eng. Soc., Vol. 29 (July/Aug.), pp. 482-491.

(5) Rosenthal, Robert. 1998. "Covert Communication in Classrooms, Clinics, and Courtrooms," Eye on Psi Chi. Vol. 3, No. 1, pp. 18-22.

(6) Lipshitz, Stanley. 1984. "The Digital Challenge: A Report" BAS Speaker Aug.-Sept., available at http://www.bostonaudiosociety.org/bas_speaker/abx_testing2.htm.

(7) Clark, David. 1982. "High-Resolution Subjective Testing Using a Double-Blind Comparator", J. Audio Eng. Soc., Vol. 30 (May), pp. 330-338.

(8) See, for example, Grant, Doug and Wurcer, Scott. "Avoiding Passive Component Pitfalls," Analog Devices Application Note AN-348.

 Free download from www.linearaudio.net © Linear Audio and Stuart Yaniger