Bridges
Mathematical Connections
in Art, Music, and Science

Subjective Fidelity and Problems of Measurement in Audio Reproduction

John V.C. Nye
Department of Economics
Washington University in St. Louis
St. Louis, MO 63130
E-mail: nye@wuecon.wustl.edu

It has been almost twenty years since Sony first advertised the brand new audio compact disc with the slogan, "Perfect Sound Forever." That slogan signified the beliefs, then widespread in the commercial recording industry, that 1) the newly adopted Redbook compact disk standard, with recordings encoded with 16 bits, sampled over 44,000 times per second, was de facto perfect, i.e. when properly done was indistinguishable from the live sound, that 2) older analog recordings (on tape) being less than perfect could be simply transferred to digital in sonically transparent recordings, and 3) that since the process of encoding was done digitally, digital copies, including those typical of making commercial discs from the master tape were absolutely identical in sound to the originals.

There were a few skeptics within the professional audio community who noted that the specifications for the new medium, impressive though they were, did not literally correspond to perfection. A still smaller group suggested that research on digital encoding made the setting of a permanent standard premature, but on the whole, opposition to the new digital medium was confined to hobbyist audiophiles devoted to the highest standards of sonic fidelity in reproduction. The early complaints, particularly those relating to the earliest digital recordings as well as the first commercial transfers from analog tape to CD, were dismissed as being the ravings of a lunatic fringe. When the problems in the transfers were clearly audible, as was often the case with some recordings, mainstream professionals sometimes responded to the complaints with claims that the flaws either a) lay in the original master tapes which were now made audible through the superior medium of digitization or b) resulted from the unfamiliarity of producers and recording engineers with the new, superior technology.

Still, complaints persisted. Complaints were voiced not just by inexperienced hobbyists but by many professional recording engineers whose livelihoods depended on producing the best recordings possible. Complaints of harsh string sound and unpleasantly grating treble were commonplace. Others complained of loss of spatial and depth information in the digital processing. More disturbing yet were claims that not all CD players sounded identical. And most troubling of all, were increasing reports in the field from professionals that compact disks mastered at different factories from the identical master tapes or from direct copies thereof did not sound alike.

While some of the complaints may have been illusory or arbitrary, objective evidence began to accumulate that the digital universe was not as simple as some of its supporters had made out in the late 1970s and early 1980s. While this is not meant to be a survey of changes in digital technology and in audio science, I will simply note some of the discoveries that have emerged in the past decade and a half. For starters, the filtering technology used in the process of reconstructing recordings converted to ones and zeroes, particularly the use of "brick wall" filters turned out to have audible and measurable defects. The analog to digital technology itself was not as settled as seemed to be the case, and there has been much room for substantial improvements in the development of professional encoders. Jitter, or timing error in the decoding of digital signals, was a more serious problem than had hitherto been suspected and it was possible to create players with audible and measurable jitter artifacts. More significantly it now seems to be the case that it is possible that timing problems at the moment of recording may be a problem and may be relevant in transferring recordings from one medium to another. The sixteen bit standard can apparently be improved with 20 bit and even 24 bit recordings. Furthermore, 16 bit recordings with a theoretical dynamic range of about 90 db (as opposed to 50 to 60 db for professional reel-to-reel tape) were seriously affected by treatment of the least significant bit. In particular, the addition of low level random noise -- called dither -- not only affects the quality of the recorded sound, but can enable 16 bit systems to resolve detail smaller than the least-significant bit. And there seems to be a small but growing body of evidence that the inability of most people to hear above 20 Khz does not rule out audible improvement by recording at higher sampling rates. While the 44.1 K standard should guarantee flat reproduced frequency response to past 20 Khz, early research seems to indicate that subtle improvement is audible by persons with average hearing who listen to recordings made with newer technologies at sampling rates of 88 or 96 Khz and higher. This has led to the development of professional analog/digital recorders, commercial CD players, and commercial recordings, which are all audibly superior to the average equivalent in 1982. Most of this information was unknown, or at best dimly perceived, when the original digital standard was carved in stone nearly two decades ago.

The bigger irony is that an increasingly large body of the most discriminating listeners, both amateurs and professionals, have rejected a narrow view of accuracy in sound reproduction in favor of a more subjective, taste based approach to listening. This is most clearly evidenced by the survival of the long-playing vinyl disc and mechanical turntable nearly two decades after the advent of the CD. While the vinyl record will never again be a mass market medium, given the cheapness with which CDs are produced and their convenience in use, vinyl lps survive both in the rarefied world of high-end audio and in the low tech world of garage band rock. Famous recordings from the late 1950s and early 1960s -- often referred to as the Golden Age of Stereo -- are now reissued in state of the art vinyl pressings and sell briskly at $30 each in specialist catalogs. Makers of ultra high-quality turntables -- those selling for upwards of several thousand dollars find themselves doing more business now than at the end of the 1980s. Also, many record producers continue to use analog tape in their professional work, and many of those who have switched to digital only do so with highly modified equipment designed to vastly exceed the performance of the "perfect" equipment of the early 1980s. Indeed, if the old CD standard leads to replacement of the compact disk by DVD, records may yet outlast CDs.

In a meeting held concurrently with the Fall 1997 Audio Engineering Society Conference in New York, a number of industry professionals came to discuss the robust state of analog recording in the studio. According to Steve Smith of Quantegy, roughly 85% of the discs listed at number one in the Billboard charts were mastered with analog tape, only 15% were on digital [1]. Bob Ludwig, a noted mastering engineer, has observed that the use of analog tape vs. digital is up substantially compared to the situation in the late 1980s. He added, "In every case, we do listening comparisons, use our ears. In not every case does the analog win. But in probably 90% of the cases where there is a shoot-out with DAT, the analog does win" [1, p. 63].

In his keynote speech to the 103rd Convention of the Audio Engineering Society held in September of 1997, George Massenburg noted,

> Thirty years into the Analog to Digital technology break (one of the longest, if not THE longest, transitions of its kind in Industrial Age history), questions persist among many in our industry (although admittedly not a majority) about digital audio, its application and its shortcomings.
>
> A quick survey of quite a few major American and British mastering houses ... reveals that more than 50% of the final masters to be transferred to CD master comes in on ½" "analog" tape [4].

How many of these trends reflect a fashion for nostalgia, and how many will instead point to the way to future engineering advances is a subject for another day. Instead, I wish to focus on a few methodological problems that seem to have arisen from a number of conceptual problems that have been entangled with the problems of subjective valuation so central to much work in the performing arts.

I want to discuss the problems of mapping the more easily quantifiable measures of engineering accuracy onto the dimension of subjective fidelity. That is, how do we get from what we can objectively measure in the lab to what we perceive and prefer in the field? It is my claim that the rhetoric of measurement, and the practical problems of mapping a multidimensional scale onto a single metric of perceived accuracy are further exacerbated by the common and theoretically mistaken practice of confusing measured importance with subjective value. The audio recording industry is a perfect backdrop for this discussion because it brings together the realms of science, engineering, commerce, and art in a strange and rapidly changing mix that is often difficult to evaluate and where conflicting claims often stem from discussion that crosses over multiple levels. I will also briefly deal with the problems of existing blind tests and their interpretation by the audio engineering community.

Measurement and Subjective Valuation

The task confronting scientists studying a phenomenon like the reproduction of audio seems straightforward: Measure as many of the parameters as possible, establish repeatable effects so that measures are reliable, relate those measured parameters to perceived differences, evaluate the mechanisms of causation for the purposes of studying the science and the engineering of reproduced music. Commercial firms take this one step further, attempting to use this knowledge for the purpose of improving recordings or designing equipment that will sell in the marketplace.

The difficulty lies in taking those measures and mapping them onto a separate dimension (or dimensions) corresponding to human judgments about which systems, or recordings sound better. But right away that confronts us with a real problem. Whose judgment counts? And more important how do we weight the variety of judgments that we observe?

To take one prominent and controversial example: Repeated tests of listener preference using a variety of groups in blind listening sessions indicate that these groups on the average rate speakers superior which have relatively flat frequency response. Put another way, the measured effect of deviations from flat frequency response are easily observed, even with untrained listeners. The effect is pronounced especially in comparison to subtler distortions having to do with matters such as pitch or phase. If the research that was performed was meticulous and exhaustive it is safe to conclude that frequency response deviations are the most easily observed aspects of high fidelity. Some engineers, particularly in their informal pronouncements, conclude -- erroneously -- that therefore, flat frequency response is the most important component of speaker design.

Notice the subtle but clear slip. The characteristic with the highest measured effect on subjective fidelity is presumed to be the most important aspect of speaker design. This begs an important question: Why should we presume that the characteristics that are most easily and most reliably observed by an average or representative listener (whatever such a creature looks like) have the greatest subjective value?

The goal of high fidelity, particularly as it relates to stereophonic reproduction deals with illusion. The ear is tricked into hearing sound that seems to have a solid presence in space despite the fact that sound emanates discretely from two separate sources. Hence stereo phonic, or "solid sound." Who is to say, however, which characteristics of objective accuracy have the most bearing on subjective accuracy? Early experiments in the 1950s indicated that in many instances listeners preferred systems with truncated frequency response to some systems with flatter frequency response [6]. Although Olson claims that this was due to the increased presence of high and low frequency distortion, the issue of non-monotonic preferences was not pursued. The issue is complicated further when two systems are not uniformly comparable and no one system completely dominates the other.

Many would say that an amplifier with uniformly low distortion measurements is more "accurate" than another amplifier with much higher average distortion measurements. However, it is well known in the literature that the nature of distortion is relevant to the perception of naturalness and accuracy. Most listeners prefer systems like those based on vacuum tubes with larger amounts of even order harmonic distortion to many transistor based systems with objectively better measurements but unacceptable levels of odd order harmonic distortion. When transistor amplifiers were introduced in the 1960s, many heralded their arrival for the ease with which amplifiers could be produced with low distortion measures and high output. Their cheapness also contributed to their dominance in the consumer market. But in hindsight, the amazing fact is that informed connoisseurs of sound, when given a choice today assign virtually no value to any transistor designs from the sixties while placing a high value on the best vacuum tube designs from the sixties. Such is the judgment of the market. It is possible that such choices are not simply faddish nostalgia, but a realistic preference for the superior sound of those tube-based designs. Indeed, work in the amplifier field since then has been heavily concerned with extending the advantages of transistors while mitigating the problems (some of which are now measurable) of the earliest designs.[1]

Other tradeoffs abound, not the least of which is the problem of the least objectionable flaw. Consider the problem of hiss from cassette or tape. Tape hiss is obvious and noticeable to even the most untrained listener. Indeed, some people prefer all digitally recorded CDs to those based on recordings transferred from reel-to-reel tape, because they find tape hiss unacceptable.

[1] Hamm, [2] claims that vacuum tube designs are often superior to solid state designs because recordings drive systems into frequent, transient overload. His hypothesis is that while all modern designs do well in dealing with steady state, undemanding loads, in practice even the most powerful amplifier must take some brief overload into account. Tube designs have more benign distortion characteristics in overload than do solid state designs, hence their generally superior subjective performance despite their inferior accuracy on standard distortion measures.

However, many experienced listeners not only can become inured to tape hiss, but find it in the long run much more acceptable than other subtler, less easily perceived flaws. These would include pitch accuracy, tonal balance, or high frequency clarity. Collectors of old 78 recordings are familiar with the tendency of some producers to so crudely filter noisy old recordings in making the transfer to CD, that the gains in background silence are offset by the losses in high frequency and ambient information.

Similarly, a musician might find tape hiss unobjectionable, may indeed fail to notice it at all after a bit, while becoming extremely disturbed by a piano recording that exhibits pitch inaccuracies or that is bleached of harmonic content.

In this instance what we observe is that in moving from an imperfect reproduction or system to another imperfect system, it is not clear which directions constitute positive improvement. This phenomenon is not unknown in mathematics, and in economics goes by the notion of the problem of the second best. In the economics literature, this problem is best illustrated by the example of an economy with severe price distortions, tariffs and subsidies, so as to be far from the theoretical ideal. One's intuition is that improvement in one or more dimensions would move us closer to the desired theoretical ideal, but as can be easily demonstrated, no such improvement is guaranteed, even if no reduction on any single dimension is observed [3].

The idea behind this observation carries over quite easily to the art and science of audio reproduction. Assume that we can perfectly characterize all the individual dimensions of a recording (a dubious proposition in my view). Assume further that in observing a specific recording or recording technology, we have established a way to improve one characteristic without altering characteristics on other dimensions. Then -- contrary to most people's intuition -- there is no guarantee that the net improvement will be positive. Clear improvements in any subset of dimensions do not map onto monotonic improvements in the subjective perception of accuracy.

This is especially troubling for researchers in fields where taste and perception are a large component of the mix. Because the problem of the second best is so debilitating, it is unsurprising that people would choose to proceed as if serious dimensional improvements were on average net improvements. But I submit that anyone interested in applying the science of audio to improving and understanding the experience of enjoying reproduced music will be doomed to frustration if he doesn't take the problem of the second best seriously. In fact, I would assert that the problem of the second best is likely to be a more severe handicap in the audio world than it is in economics.

Because such a notion is counterintuitive, it is difficult to build such ideas into systematic investigation of the science of sound reproduction. There are any number of phenomena which, if considered thoroughly, imply a substantial rethinking of the implicit model of human hearing used by many professionals. These include, not just the state-of-the-art work in digital encoding, but (to take a single, striking example) also the extension into audio of the problem of stochastic resonance.

Most models of sound reproduction assume that there is a monotonic relationship between the signal to noise ratio and the clarity or intelligibility of sound. Work on distortion measures assume that less measurable distortion is always good with asymptotically small amounts of distortion being inaudible in steady state tests. But work on stochastic resonance begun in the early 1980s indicates that under many circumstances the human ear can actually perceive many "inaudible" sounds more readily when a small amount of noise is added to the mix (For a short introduction see [7]). This work has potential for improving as well as complicating audio research, because it attacks one of the bases for standard audio engineering -- the use of listener discrimination of steady state test tones or other test signals as a basis for drawing conclusions about reactions to more complex signals. If there is no straightforward way to combine known work on steady state signals to accurately predict user reaction to complex music, then much of the work on measurement, or at least its simple interpretation, becomes open to question. This is obviously an issue that specialists in engineering and physics as well as psychoacoustics are better qualified to pursue.

Placebo Effects and the Choice of a Loss Function

But the overriding problem of an overly narrow interpretation of existing measurements is its implicit appeals to valuation, which are and must always be subjective.

Many critics have disparaged the heavy reliance by top record producers, recording engineers, and audio magazine reviewers on subjective listening impressions on the grounds that humans cannot hear reliably.

It is further observed that many people suffer from a placebo effect that disappears in double-blind testing, and that in general people are too unduly influenced by uncontrolled factors that make for difficult scientific verification. Indeed, work by Floyd Toole and others does confirm that people's auditory judgments are influenced by what they see as well as hear. But it is a long logical leap to go from saying that people can be fooled to saying that everything they hear that can't be demonstrated in a blind test is illusory.

The standard assertion of placebo effects, and the related claims about the comprehensiveness of existing test protocols mixes the problem of reliability of human observation with the value and importance to the listener of unquantifiable anecdotal observation. That human beings do not hear in ways easily consistent with standard experiments in no way invalidates the virtues of human perception. But, at the most extreme, some adherents of controlled double blind testing do not admit any listening impressions not derived from strict testing procedures any standing as data.[2]

Though controlled repeatability is the hallmark of some of the best work in the natural sciences, and is the primary form of testing in the medical literature, it is not the be all and end all of good science. Geology and astronomy permit observation based on uncontrolled experiments as do almost all the social sciences. Second, the double-blind experiments that have been performed have not always been identical in condition to those under which people listen to their equipment. This compounds the standard statistical problem that it is very difficult to prove a negative. If we run a series of tests that show no statistically significant difference, is it because a) no differences were heard b) some differences were heard but did not aggregate up to statistical significance, or c) only a minority could hear the differences? Further, even if a) were true, could the lack of differences have been due to poor test setup, poor listening training, or

[2] When dealing with one "scientifically inclined" speaker designer I was astonished to find that he was unwilling to voice an opinion of the sound of a system using his own speakers because "He wasn't very good at open loop listening." When told by a noted record producer that the setup was too bright and something needed to be adjusted he averred that these weren't controlled conditions and he didn't know whether the producer was right or not without being able to take measurements. This may be an absurd example, but it is not atypical of the extremes that blind reliance on the "science" of audio may lead to.

lack of an ability to isolate which differences could be heard in such a way as to guarantee repeatable audibility? Often so much effort is made to control for the possibility of erroneous differences that test sensitivity is compromised. In statisticians' jargon, Type 1 error, or false positives, are minimized in the test protocols, but Type 2 error, or false negatives are not.

Still worse and quite common, is the blind reliance on classical statistical criteria. Most journals use the 95% confidence interval as indicating significance, but this is purely arbitrary and often wrongheaded [5]. By the very nature of modern science, we minimize false positives. Nonetheless, while it may make sense for us to be extremely conservative in calling a difference significant for the purposes of our science, such conservatism may not and often should not apply to individual listeners when the cost of making mistakes is small.

Imagine the following scenario: Engineer A is recording an important jazz band. He finds that machine X sounds better than machine Y, but when challenged to a blind test by engineer B, he finds that his capacity to select X over Y is only significant at the 60% level, far below standard significance levels. However, the jazz band that engineer A records is extremely important and critical of sound quality. Their performances are widely regarded and may be of historic significance. Assuming that the cost difference between machines X and Y are not outside his budget, should the engineer stick to the cheaper machine Y, or should he use X, believing it to be better but knowing there is a large chance that his perception of its superiority is completely illusory? I submit, that no reasonable individual would encourage the engineer to choose machine Y.

Even if we grant the real possibility that a large fraction of perceived psychoacoustic differences are illusory, it still makes sense for most individuals to follow their perceptions if the costs of doing so are small. Once one concedes this, the question of measurable differences takes on new dimensions. Modern science, particularly in engineering or medicine, proceeds on the assumption that no effect should be claimed unless the probability that we are mistaken is extremely low. But this is the wrong attitude in a subject where the benefits of uncertainty may be high, and the cost of mistakes small.

Furthermore, the notion of a small cost is an entirely subjective one, as is clear from the literature in economics. For some a hundred dollars may be a lot. For a billionaire, or a dedicated musician, or a record producer whose work sells in the hundreds of thousands of copies, paying $100,000 for a small perceived difference may be a bargain.

The use of controlled double-blind testing takes its cue from work in medicine and psychology and is inappropriately applied to work in audio engineering. It is worth reviewing the justification for testing protocols in medicine and noting how applications in different fields require value judgments that are not in any definitive sense, objective or "scientific." The standard argument in the medical literature combines a reasonable testing procedure with an implicit set of value judgments. In particular, the fear of unwelcome side effects from improperly tested or approved drugs leads to an extremely cautious and conservative approach on the part of the medical community. Where there is doubt as to the efficacy of a new drug, the entire methodology of strict controls and large-scale randomized testing is designed to minimize the probability of approving a drug that is not in fact, efficacious, or is less than already existing substitutes. Such a procedure purposely permits the possibility that the tests may lead us to deny approval of an effective treatment, if such treatment does not show clear benefits in controlled tests.

Such an attitude is especially useful in a realm where unexpected side effects are costly in terms of human suffering, and potentially dangerous from the standpoint of litigation. It is not at all clear that such a policy makes sense when the alternatives are innocuous -- why shouldn't we approve a cheap drug with no side effects even if it might be useless -- or as in the case of terminally ill patients suffering from incurable diseases, when there is little to lose by trying half-baked technologies.

It seems foolish to adopt such conservative standards in fields like audio engineering, especially when the statistical bases of the tests themselves are so shaky. Indeed, it may be foolish to apply such strict standards in any fields. As McCloskey has emphasized, the use of statistical tests -- notably the designation of the 95% confidence interval as "significant" -- has tended to be abused in the medical, economic, and psychological literature confusing statistical and substantive significance [5]. Bayesian statisticians would further cavil with the almost flagrant use of strict statistical criteria for judging significance. Without a loss function, without some sense of what the cost (in the broadest sense of the term) of an incorrect judgment is, the heavy reliance on "significance" may not be the right model for audio engineering. In the medical literature, there is already some question as to whether rejecting or accepting articles simply on the basis of significant statistics is a reasonable procedure. One might argue that an arbitrary but conservative convention is sensible in a field where human lives are at stake. But when considering the audibility or quality of sonic reproduction, *surely the cost of dismissing an important phenomenon that might be real should be balanced against the standard problem of accepting as true, an effect that is merely illusory.*

I have never understood why the standard should be otherwise. Performing artists know all too well that sound matters, and that any number of small details affect the quality of a performance and how it is perceived. They may not be able to articulate why a difference is heard. They may not hear a difference reliably, or may give incorrect reasons for what they hear, or even hear things which are not there, but in the final analysis only human valuation really counts. Scientific research may give one clues as to why, for example, a Stradivarius or Guarnerius sounds better than a modern copy. A great violinist however, will dismiss as nonsense any claim that a modern violin sounds better than a Strad simply because it measures as well, and simply because performers cannot prove that a Strad sounds better.

So many of these problems reside in the unstated assumptions and judgments researchers make about what are real or important differences. It is particularly striking to see these methodological errors repeated even when authors are aware of how erroneous certain postures have been in the past. In a perceptive article recently published in *Audio* magazine, J. Robert Stuart, Chairman and Technical Director of Meridian Audio, a highly respected digital equipment manufacturer, detailed some of the many breakthroughs that have been made in the last decade about how to improve and properly evaluate the existing digital standards, with a special emphasis on the subtle problem of dither implementation and its real, measurable, and audible effects on sound. Yet, having detailed how far we have come, and how little we knew at the start of the digital era, he concludes the essay by proposing a fixed set of standards derived from existing experimental research and arguing that, "we should exceed these requirements *only* when there is no detrimental cost to doing so" [8]. Again, we have someone imposing an implicit loss function, indeed assigning an implicit value of zero to potential improvements and thus setting the bar for moving to still higher future standards at an arbitrarily high level.

The conflict between measurement-oriented professionals and the most anti-scientific audiophiles (some of whom reject *all* attempts at measurement) would seem to be a perfect case of C.P. Snow's two cultures colliding. But this ignores the large number of scientifically trained professionals who are in the same camp as the hobbyist and performing artist. Too many professionals have observed that much in the world of audio reproduction cannot easily be predicted by known measurements. The conflict arises from too narrow an interpretation of the demands of scientific methodology and an unreasonable fear of judging sound and reproduction on the basis of perceptions that may turn out to be illusory. But when dealing with protocols at the boundaries of art, science, and commerce, the real danger is not imprecision, but bad art in the service of simplified engineering.

This article has sought to make two points that are well-known in rather different contexts. Both are representative of larger problems dealt with in greater detail in the literature of economics and statistics. The first deals with the difficulty of summarizing a multi-dimensional phenomenon with a one-dimensional measure. This is often referred to as the index number problem or the problem of aggregation. To the extent that subjective fidelity can be thought of as an index of "goodness" or perceived accuracy, the problem for audio engineering is finding a proper mapping between the known, measurable characteristics of music reproduction systems and this undefined and unobservable index of subjective fidelity. The second problem concerns the difficulty of doing research that incorporates explicit or implicit value judgments in situations where the necessary loss function is either unknown or unspecified. This difficulty is more commonly discussed in the economics literature where researchers try to make the most of limited comparisons in a field where standard practice restricts our ability to compare different utility functions or pronounce judgment on differing tastes and preferences.

The latter problem is of great relevance for any applied field that makes stringent use of classical criteria of statistical significance as well as for the policy analysis of issues where matters of valuation are not dealt with as rigorously as the controls on the tests themselves. While this essay has mainly dealt with the narrow problem of audio engineering, it has also touched on the problem of significance in the medical literature where judgments about the "worth" of a drug or procedure are often discussed as if there were some universally applicable cost-benefit formulas. Similarly, these difficulties extend to any number of areas from welfare policy, to the science and politics of the environment, where the weight of scientific evidence will only go so far in serving as a guide to action. Reconciling what is known for certain, with one's tolerance for taking risks and evaluating the costs of the unknown, require a more extended dialogue and more sophisticated treatment of valuation than is typically observed.

It is not an exaggeration to suggest that this short discussion of the problems of high fidelity reproduction is relevant to any number of fields where the quest for objective, impersonal knowledge overlaps with issues of choice and opportunity cost.

References

[1] Michael Fremer, "Analog Corner," *Stereophile*, Vol. 27, no. 1, January, 1998.

[2] Russell O. Hamm, "Tubes vs. Transistors: Is There an Audible Difference?" *Journal of the Audio Engineering Society*, presented September 14, 1972, Internet paper version.

[3] Jean-Jacques Laffont, *Fundamentals of Public Economics*, MIT Press, 1990.

[4] George Massenburg, Keynote Speech to the 103rd AES Convention, 1997 as reprinted at http://www.aes.org/events/103/.

[5] D.N. McCloskey, "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests," *American Economic Review*, Vol. 75, no. 2, May, pp. 201-205, 1985.

[6] Harry F. Olson, *Music, Physics, and Engineering*, 2nd Edition, Dover Press, 1967.

[7] Ivars Peterson, "The Signal Value of Noise: Adding the Right Kind of Can Amplify a Weak Signal" *Science News*, Vol. 139, February 23, p. 127, 1991.

[8] J. Robert Stuart, "Digital Audio for the Future, Part 1," *Audio*, Vol. 82, no. 3, pp. 30-39, 1998.

[9] John Watkinson, *The Art of Digital Audio*. London and Boston, 1988.