AES Convention Papers Forum

The Audibility of Typical Digital Audio Filters in a High-Fidelity Playback System

(Subscribe to this discussion)

This paper describes listening tests investigating the audibility of various filters applied in high-resolution wideband digital playback systems. Discrimination between filtered and unfiltered signals was compared directly in the same subjects using a double-blind psychophysical test. Filter responses tested were representative of anti-alias filters used in A/D (analog-to-digital) converters or mastering processes. Further tests probed the audibility of 16-bit quantization with or without a rectangular dither. Results suggest that listeners are sensitive to the small signal alterations introduced by these filters and quantization. Two main conclusions are offered: first, there exist audible signals that cannot be encoded transparently by a standard CD; and second, an audio chain used for such experiments must be capable of high-fidelity reproduction.

Authors: Jackson, Helen M.; Capp, Michael D.; Stuart, J. Robert
Affiliation: Meridian Audio Ltd., Huntingdon, UK
AES Convention: 137 (October 2014) **Paper Number:** 9174
Publication Date: October 8, 2014
Subject: Perception: Part 2

Click to purchase paper or you can login as an AES member to see more options.

(Comment on this paper)

Comments on this paper

Stefan Heinzmann

Comment posted December 19, 2014 @ 16:26:18 UTC *(Comment permalink)*

I see a number of problems with the paper, some of which are:

The conclusions in the abstract and in the introduction differ markedly from those offered at the end of the paper. They are also not adequately supported by the research presented in the paper.

The introduction contains a lengthy series of speculations which seem to be preoccupied with casting doubt on some past research, yet the paper essentially fails to substantiate the speculations. This comes dangerously close to being unfair.

The criticism of the ABX test procedure that is offered in the introduction is poorly justified. The "cognitive load", as called by the authors, is entirely under the control of the listener in an ABX test, since the listener selects when to switch and what to switch to. There is no requirement to keep all three sounds in memory simultaneously, as criticised by the authors. Consequently, it is unclear what advantage the method chosen by the authors offers over an ABX test. Furthermore, the informal use of the term "cognitive load" seems to suggest tacitly, that a higher "load" is detrimental to the ability to distinguish between different sounds. I'm not aware of any study that confirms that. Indeed, one could just as easily suspect the opposite, namely that the availability of more sounds would increase this ability. Neither of those suggestions can of course be taken for granted. The authors shouldn't appeal to their interpretation of common sense when criticising a test method, and rely on testable evidence instead.

The quantization to 16 bit was accompanied by either no dither, or RPDF dither. As the authors rightly state, neither is satisfactory, a well-known fact. However, using nonoptimal test conditions defeats the aim of showing deficiencies of the CD-format itself. If artefacts are uncovered with this setup, they may be attributed to the test conditions rather than the format. This means that the conclusion drawn by the authors, regarding the CD-format, is unjustified by the result of their research. If the CD-format as such is under scrutiny, the aim must be to remove all other factors as much as possible. This rings especially dissonant with the authors' criticism of the work of others offered in the introduction.

While the usage of no dither or RPDF dither was justified with a reference to alleged deficiencies of *some* real-world implementations of converters, the authors chose unusually high cut-off frequencies in their lowpass filters. This can only make it more likely that any uncovered artefacts are due to the test conditions, and not due to the CD format. Again, the authors ought to make clear whether they want to address deficiencies of real world implementations, and if so direct their conclusions at those implementations, or whether they want to address deficiencies of a format, in which case they should attempt to exclude or reduce the deficiencies of

real-world implementations from their test setup as much as possible.

Hence I conclude: The research that is presented in this paper shows evidence that supports conclusions 1 and 2 at the end of the paper. I don't see how it supports conclusions 3 and 4, which appear speculative to me. Conclusion 5 isn't actually a conclusion, it rather seems to describe a preconception of the authors, which affected their design of the test procedure. I don't see any attempt on their part to investigate to what extent this preconception is actually valid.

Of the two main conclusions offered at the end of the abstract and the end of the introduction, neither are supported by the research, indeed the tests conducted were not designed to address those questions. That doesn't make the conclusions automatically false, but it casts serious doubts on the authors' interpretation of their own findings.

Kind regards

Stefan Heinzmann

(Respond to this comment)

Amir Majidimehr

Comment posted March 13, 2015 @ 16:45:09 UTC

(Comment permalink)

This is in response to Mr. Heinzmann's comments. While I think the paper could stand a bit more clarity on its positions and nature of testing, the comments below by Mr. Heinzmann can be addressed:

>>> The criticism of the ABX test procedure that is offered in the introduction is poorly justified. The "cognitive load", as called by the authors, is entirely under the control of the listener in an ABX test, since the listener selects when to switch and what to switch to. There is no requirement to keep all three sounds in memory simultaneously, as criticised by the authors.

While there may not be a requirement that is what is routinely done by listeners. The instructions for ABX tests usually specify listening to A, then B followed by X and determine if X matches A or B. Intuitively then, the user will attempt to do the same by trying to remember all three stimuli. A trained listener may not follow such instructions (see below) but such were not used in the test, nor would that represent the general public who would take such tests. They will try to listen to all three clips, and if differences are small, struggle to tell if X is a better match to A or B. Therefore the cognitive load is most definitely there and needs no proof of its existence.

>>>Consequently, it is unclear what advantage the method chosen by the authors offers over an ABX test.

Actually the test used by authors can be thought of an ABX test with better instructions. One can always listen to A, disregard B, and vote whether X matches A or not. And in doing so, have the ABX test behave exactly as was done in the paired comparisons used in this research. In that sense, both the authors' comments and Mr. Heinzmann's are moot. What was run was an ABX test albeit, optimized so that the listener was not asked to do more work than necessary.

>>>Furthermore, the informal use of the term "cognitive load" seems to suggest tacitly, that a higher "load" is detrimental to the ability to distinguish between different sounds. I'm not aware of any study that confirms that.

>>> Indeed, one could just as easily suspect the opposite, namely that the availability of more sounds would increase this ability. Neither of those suggestions can of course be taken for granted. The authors shouldn't appeal to their interpretation of common sense when criticising a test method, and rely on testable evidence instead.

The research is plainly there in the form of capacity of short-term auditory memory which is measured in seconds. Any test that spills over that capacity, will force the listener to rely on longer term memory which is far less precise. When differences get small, we must do everything we can to enable the listener to utilize short-term memory. It is for this reason for example that very short music segments are used in MPEG reference test clips for test of lossy audio codecs. More is not better at all. I welcome Mr. Heinzmann presenting his research that more is better.

>>>The quantization to 16 bit was accompanied by either no dither, or RPDF dither. As the authors rightly state, neither is satisfactory, a well-known fact.

You mean from audibility point of view? If so, that is not well-known fact or better said, remotely accepted fact by skeptics. This test for the first time demonstrates that conversions to 16 bits using rectangular dither or truncation could very well be audible in double blind controlled listening tests. Until now, such small differences were routinely considered to be inaudible. If we now take such results as well-known fact, then we have taken a large step forward in bridging the gap between the believers and non-believers in high-resolution audio.

>>>However, using nonoptimal test conditions defeats the aim of showing deficiencies of the CD-format itself. If artefacts are uncovered with this setup, they may be attributed to the test conditions rather than the format. This means that the conclusion drawn by the authors, regarding the CD-format, is unjustified by the result of their research. If the CD-format as such is under scrutiny, the aim must be to remove all other factors as much as possible. This rings especially dissonant with the authors' criticism of the work of others offered in the introduction.

Two requirements are imposed by CD format: 44.1 KHz sampling and 16-bit quantization. The former was tested by itself. 192 KHz/24-bit content was filtered down to 22.05 KHz while keeping the bit depth at 24-bits with TPDF dither. That one processing which is a mandatory pre-requisite of mastering for CD, was found to be audible with better than 95% confidence. It matters not then what the additional step of quantization to 16 bit does. The “damage” was already done in the filtering and the war of transparency lost.
Also keep in mind that the CD spec does not mandate a specific re-quantization. As such, you have no idea what conversion was used to 16 bits in music one buys. Anything from truncation to noise-shaped TPDF dither may be used and everything in between. So unless you can represent that large amount of music one buys in CD format is using optimized dither, then as a practical matter that criticism is of limited value.

The only way to be assured of transparency is to get the originally mastered stereo track prior to filtering and re-quantization. Anything else can mean some audible compromise. That is what this paper directionally shows and the high-level, take away message.

(Respond to this comment)

John Stuart

Comment posted March 16, 2015 @ 13:49:34 UTC
(Comment permalink)

This is in response to the comments from Messrs Heinzmann and Kreuger.

We would like to address them here in advance of publishing further experimental data in this area.

You are correct that the concluding remarks move beyond the abstract and, to an extent, this reflects the fact that the abstract was submitted some time ahead of the paper. As noted in the Summary (4.4), this is a report of a pilot in an ongoing study.

We do not agree with the comments relating to the introduction. The central question in this paper was to determine whether the addition of certain low-pass filters could be detected in an audio chain. We do emphasise in the introduction the necessity to ensure that the filter under test should narrow the overall system bandwidth. It seems logical that the playback system should be wideband and documented, that the signal should have known provenance, be repeatable and be of suitable quality, and that there should only be one change made in the signal path between test conditions. Any criticism of any of the 6 listening tests referred to rested only on these points or the absence of such information.

Regarding the choice of psychophysical test: we chose to use the 1AFC (one-alternative forced-choice) same-different (AX) paradigm, one of many double-blind forced-choice paradigms that are appropriate to the task in hand, that is, where the basis on which listeners discriminate the stimuli does not need to be known a priori. Other possible options included the 2AFC (two-alternative forced-choice) ABX paradigm and the 4IAX (four-interval AX) paradigm, a 2AFC version of AX, in which a listener must decide which of two stimulus pairs contains a difference. Pre-testing indicated that listeners found our test in any paradigm quite

difficult, and their feedback indicated that they preferred fewer intervals per trial due to finding the task tiring. This could have been due to the signals we used being fairly long, lasting around ten seconds each; often where ABX is used in psychophysics the stimuli are tones, noises or speech-sounds like vowels, each lasting only a few seconds or even milliseconds. 4IAX was found to be nearly unusable for this task.

It is not uncommon to believe, as we do, that the ABX test is “hard” for listeners, and hence possibly sacrifices some sensitivity and reliability over simpler tasks such as same-different (AX, see Lass 1984, Crowder 1982). For example, Pisoni (1975) compared results from ABX and 4IAX in the same subjects for short speech-stimuli, and found that the 4IAX invariably gave smaller threshold estimates. However, we accept that the use of the term “cognitive load” was perhaps over-reaching as we used it.

One problem that exists with the 1AFC version of the AX test, as we employed, is that of potential bias in the results due to the internal decision criterion of a particular listener. Although in this paper we grouped the subjects together and analysed performance using the binomial distribution, in the future we will consider adding analysis methods from signal detection theory (analysing hits, misses, correct rejections and false alarms) to measure and adjust for this. This was not possible for the data collected here as the sample size was not large enough.

The empirical observation remains that our listeners could discriminate filtered from unfiltered signals at a level above chance for five out of six conditions (with a risk of Type I errors of 1 in 20) and that significant ($p<0.05$) effects of all parameters were found in the results for the high-yield sections. We find it hard to see how these observations could arise from any other interpretation than sensitivity to our signal processing. As our experiments progress, we will be interested to see if our tentative conclusions are supported by new data.

Turning to the comments on dither: we know that in order to approach transparency TPDF is the minimum that should be accepted. A quick search of the AES library shows three papers by Stuart (one of the present authors) on this very topic, including the use of noise-shaping. Our paper states that the core test carried out here, namely the introduction of a filter into a 192-kHz 24-bit channel, used TPDF at the LSB in the filter.

As stated in the paper, we added the 16-bit quantisation as a probe, mostly out of curiosity, because we were aware that certain converters have used sub-optimal dither in their multistage filter chains in an attempt to preserve signal/noise ratio. The quantisation and dither tests were reported for information but are not central to the point of the paper. The sampling process traditionally requires bandwidth limiting, instantaneous sampling and quantisation. We aimed to determine if the first step alone could be detected, although we reported and commented on the quantisation. We did not set out to examine the CD format as such; however the fact that a band-limiting filter at 22 kHz was detectable in the 24-bit context should give some pause for thought. The choice of 192-kHz content was to ensure that the band-edge of filters used in the recording and for reconstruction in the DAC was sufficiently removed from that of the test filter.

Regarding the specific criticisms of the conclusions: Point 1, 2 and 4 (some segments of the music made the filter easier or harder to detect) are supported by our findings, and 4 particularly by reports of listeners telling us what they listened for. Point 3 is perhaps worded unhelpfully generally, but it is not untrue that our results are consistent with such a temporal smearing hypothesis; we do not claim that our results support this hypothesis. Point 5 was intended to lead to further work, and the criticism addressed above regarding the use of the term “cognitive load” is acknowledged.

As stated earlier, we have continued this series of experiments using different filters (including both shorter and minimum-phase designs) and will be reporting these findings in the near future.

Dr Helen Jackson, Dr Michael Capp and Bob Stuart

References:

Crowder, R.G. (1982) A common basis for auditory sensory storage in perception and immediate memory. Perception & Psychophysics. Sep;31(5). p477-483. doi:10.3758/BF03204857.

Pisoni, D.B. (1975) Auditory short-term memory and vowel perception. Memory & Cognition. Jan;3(1). p7-18. doi:10.3758/BF03198202.

Lass, N.J. (ed.) (1982) Speech and Language: Advances in Basic Research and Practice, Vol. 10. London: Academic Press Inc.

(Respond to this comment)

Stefan Heinzmann

Comment posted March 27, 2015 @ 19:56:12 UTC

(Comment permalink)

Thank you, Mr. Stuart, for the clarifications. Let me add one of my own before going into the details: I didn't intend to say that the conclusions of the paper go beyond those in the abstract. They are simply different, and if anything, I would tend to say the opposite, namely that the conclusions in the abstract go beyond those in the paper. My main criticism, however, is that they are not adequately supported by the research presented. I am looking forward to seeing your further research that you say will close this gap.

Apart from this, there are two main topics which I would like to address in turn. The first is the criticism aimed at the ABX test method, and the second is your choice of filter charateristics.

I was under the misapprehension, that you were criticising the ABX test method as used in more recent times. The answer by Mr. Krueger, and your choice of references that you provided with your answer, makes it very plausible that you are actually criticising a form of ABX test where the A, B and X stimuli are presented once in this order, and the listener, who has no influence on the test, is then asked whether X was A or X was B. In this case, it is understandable why you are concerned about the strain on the listener. Here, it is indeed necessary for the listener to remember the sounds in order to compare them.

I was not aware that this primitive form of ABX testing was still being used widely, particularly when trying to identify subtle differences. Improved ABX testing procedures and corresponding hardware support have been known and used for decades, which allow the listener to switch at will between A,B and X any number of times and at any point in time, and indeed your own test method allowed for the same, except of course for the lack of a stimulus B. It was your discussion of the Meyer/Moran experiment in particular, which led me to believe that you were actually criticising their way of doing ABX. Not so, as I realize by now.

It does, however, beg the question why you didn't simply resort to a more modern form of ABX, which doesn't have the problems you suspect, instead of dismissing it entirely. In any case, the question whether "modern" ABX is inferior to other approaches, such as yours, remains unanswered, whilst the criticism you have aimed at ABX has been addressed a long time ago by introducing ABX switching hardware operated by the listener.

Regarding the second topic, namely the choice of filter characteristics, I have to support Mr. Krueger. I tried to find A/D converter chips amongst my collection of data sheets, which offered a transition band as narrow as the one you used for your experiment. Apart from a chip by ESS which had a freely programmable decimator, I only encountered wider transition bands, even when the chips offered several choices. It was only a cursory look, perhaps a more thorough search would have uncovered some more examples, but I don't understand how you come to your opinion that your choice represents a typical situation encountered in the field. My own perception of the market has been for quite some time now, that the transition bands have become wider, sometimes beyond the point where I would find the risk of aliasing effects to be justifiable. So my fear is that the market is more likely to err on the side of too wide a transition band.

Your research is of course valuable in showing that too narrow a transition band may have a negative effect, too. If this leads to a realization of what transition band is "right" for which given sampling rate, it can only advance the state of the art. My own feeling, however, is that the existing converter chips are in their majority already quite close to this best choice. Yet I would find it most welcome to investigate the root cause of the differences that your experiment found audible. You offer some hypotheses that would need substantiation.

I still believe that you are making way too much of your findings. It

is far from clear that your result can be seen as pointing towards a deficiency of the CD format. If you weren't implying to judge the format as such anyway, as you say, your wording of the abstract and of the introduction was certainly unhelpful. This is also evidenced by the public reaction it has attracted. I hope that this can and will be put right in the upcoming episodes.

Kind regards

Stefan Heinzmann

(Respond to this comment)

Arnold Krueger

Comment posted February 23, 2015 @ 15:20:18 UTC *(Comment permalink)*

I have a problem with this paper's description of the ABX test, which seems to be based on the classic but irrelevant 1950 Munson and Gardiner JASA paper rather than the more recent and relevant 1982 Clark JAES paper.

I agree with Stefan Heinzmann's comments above about the use of either no dither or RPDF dither rather than the industry standard TPDF dither.

It appears that the dither used was spectrally unshaped, whlie it has long been known (for example as expounded upon in the JAES by Vanderkooy and Lipshitz, etc.) that for critical applications perceptually shaped dither should be used.

My studies of modern 44.1 KHz DACs suggest that transition bands on the order of 2 KHz are common and that the ca. 500 Hz transition bands used in the simulations are atypical.

The sample rate of the simulated digital filters was apparently 192 Khz, but in fact typical digital filters used in modern DACs run at 8x or higher or 352.8 Khz.

In my mind the above points don't exactly support the phrase **"Typical Digital Audio Filters in a High-Fidelity Playback System"** used in the title.

(Respond to this comment)

Subscribe to this discussion

To be notified of new comments on this paper you can subscribe to this RSS feed. Forum users should login to see additional options.

Join this discussion!

If you would like to contribute to the discussion about this paper and are an AES member then you can login here:

Username: dekkersj

Password:

Login

If you are not yet an AES member and have something important to say about this paper then we urge you to join the AES today and make your voice heard. You can join online today by clicking here.

©2015 Audio Engineering Society - All Rights Reserved