Text of paper presented at the 125th AES Convention, October 3, 2008

Detecting Changes in Audio Signals by Digital Differencing

Bill Waslo,

Liberty Instruments Inc., Liberty Township, OH 45044 USA
bwaslo@gmail.com

ABSTRACT

A software application has been developed to provide an accessible method, based on signal subtraction, to determine whether or not an audio signal may have been perceptibly changed by passing through components, cables, or similar processes or treatments. The goals of the program, the capabilities required of it, its effectiveness and the algorithms it uses are described. The program is made freely available for use in such tests.

1. INTRODUCTION AND BACKGROUND

1.1. The Audibility of Signal Differences

Throughout the history of audio engineering and (particularly) of audio component marketing, a number of discoveries have been announced for new types of distortions. Many factors have been said to cause or correct problems and to have noticeable audible effects in sound reproduction systems. New product concepts are commonly promoted as cures for such ills. Some examples include use of special cable geometries, amplifiers with particularly high slew rates, chemical treatments for CD disks, or devices intended to control electromagnetic interference. In fact, according to some audiophiles, nearly anything in or even near a high-resolution audio system can affect its sound.

But while such claims are common, objective evidence seldom can be found showing that these claimed distortions or factors can actually be differentiated by only hearing sound. There is often significant skepticism about whether some of these things really can affect an audio signal at *all*, much less to any audible extent. Many testimonial descriptions exist, but there is rather little that can be repeatably demonstrated. Even should a researcher choose to accept that such a claimed effect *might* be real, he would have no certain way during a product development to know whether or not he is improving related performance.

Objective testing methods for audible effects, such as double-blind A/B or ABX, do exist and are capable of verifying audibility of some changes in an audio signal or system [1]. But these methods can be time-consuming and expensive to implement rigorously, and while they can confirm an effect to be audible, they can never conclusively prove any one factor to *not* be audible. A negative (inaudible) result can at best conclude that audibility wasn't demonstrated under the particular given conditions of the test. And should results strongly imply that an effect can not detected by ear, that conclusion is likely to be routinely dismissed by much of the high-end audiophile community. The other components in the system are accused of lacking adequate resolution to preserve subtle changes, or listening conditions during the test may be thought overly stressful or otherwise atypical. Switch boxes used in the tests are suspected of degrading audio performance and masking the differences being listened for. For these audiophiles, believable conclusions are achieved only through "sighted" listening tests in which the listener already knows what he is listening to at each moment, describes the sound "subjectively" and (at least consciously) trusts only his ears. Such results, though, are of little or no use in engineering developments or scientific research.

1.2. Resolving Differences by Signal Subtraction

While there may not be a way to prove a claimed effect is inaudible, it is possible at least in principle to determine whether any two audio signals are actually different. By combining analog signals in opposite polarity, or by subtracting appropriately aligned digital samples, an audio recording of just what is different between the two recordings can be made. If this recording is silent, then it is reasonable to conclude that there is no difference between the two signals to *be* heard. If those two original signals resulted from including and not including some tested product, then a silent difference recording indicates that whatever was being tested actually had no audible effect. Analog versions of similar tests have been devised by Baxandall, Hafler and other researchers [1, 2, 3].

In a paper presented in 1991, Dunn and Hawksford [4] described a differencing system that utilized recorded digital signals made at the input and the output of a device such as an amplifier. The setup was stimulated by program material provided by a CD player with the recording clocks locked to the CD player's sample clock. The CD player provided repeatability and the use of a common clock maintained synchronization. The recorded signals could then be processed afterwards to remove expected effects through linear response equalization, time shifting or gain adjustment before extracting the difference information between them to reveal any distortions caused by the tested device. This approach greatly simplifies the requirements for difference tests, freeing the researcher from timing constraints (both signals need not be "live" simultaneously), while allowing the results to be re-examined aurally even long after the test was performed.

The Windows based software program discussed in this paper is based on Dunn's and Hawksford's method. Since that original paper was written, inexpensive yet high quality recording soundcards have become available for computers, simplifying the hardware requirements for the test and enabling its use over a broad range of sample rates and at higher bit resolutions. This program, called "Audio DiffMaker", has been released to the public domain and is intended to take advantage of this hardware for difference testing.

Use of such subtractive methods transforms the testing for sound changes from a task of hearing whether two sounds are different to the much simpler one of merely discerning whether or not the isolated difference recording can be heard. Of course, achieving an audible isolated difference does not prove that it will also be audible when accompanied by those louder parts that the original recordings have in common. But the availability of the recorded difference can provide a convincing piece of evidence that what is being tested is or isn't having any unexpected effects on the audio signal. The result can be evaluated numerically (for instance, by evaluating the relative energy in the difference signal relative to the original signal) or (perhaps less objectively), by simply listening. Evidence in the form of a WAV file can be more accessible to a listener than a data point or a graph – he is in fact encouraged to still "trust his ears".

2. CLASSIFICATION OF DIFFERENCES

Both an advantage and disadvantage of the differencing method is that it is extremely sensitive to very small changes. At times it can be challenging to obtain a deep null or quiet difference track even with two recordings made one after the other under otherwise identical conditions. When the recorded signals have been processed through any analog circuitry, there will always be some noise energy left in the resulting difference recordings should the playback gain be increased enough. And though not all possible differences are of equal interest, even tiny amounts of any can prevent achievement of what might otherwise have been a silent difference track. The unwanted differences must be avoided or compensated if others are to be either reliably exposed or shown to be effectively absent.

There are a number of “uninteresting” differences that can appear. Gain differences are easily explainable and curable, as are channel imbalances in multi-channel systems. Variations in the starting points of recordings relative to the beginning of a track or differences in the lengths of recorded tracks can be expected to result from the experimenter’s timing in clicking the Record or Stop buttons. Similarly, polarity reversals of signals are easily cured and while worth noting, are not of interest in terms of these tests. The software needs to recognize and deal with these factors so that they do not dominate the result.

Some differences might be classified as “mostly uninteresting”. Under some circumstances (such as when they are unexpected) these might represent a desired result, but provision should be made to deal with them in other cases. Minor speed or rate differences on the order of a few parts per million might indicate normal drift in clock oscillator circuits. Linear frequency response changes will often be expected but usually are not of particular interest and can be more easily determined using other techniques. Dunn and Hawksford used an additional MLS based frequency response measurement made of the setup to equalize such linear distortion effects when expected. The Audio DiffMaker software provides a similar capability using a log-swept sine wave stimulus to measure the linear frequency response characteristic (and impulse response) of both signal producing situations. Using the log-swept sine wave stimulus has the advantages over MLS of a generally higher dynamic range as well as the ability to reject effects of harmonic distortions in the measured impulse response [5].

Noise (including hum, spurious tones, and EMI) is a potentially troublesome factor, as it will appear in the difference tracks and can hinder the program’s ability to compensate the other effects described above, often leaving low level error residuals of the original signals. The nature of the differencing process is that any process that is applied imperfectly will leave some residual signal, while deep nulling requires that everything goes well and that both signals were indeed effectively the same. The significance of any residual energy that remains is subject to judgement. For most people listening to the difference tracks, a weak audible hiss will sound totally unlike the program and might be ignored. In cases where noise is strong enough to prevent proper program operation or is high enough in level that it might mask other differences, it can render a test inconclusive.

The last class of differences includes the “interesting” ones. There could be a new or unexpected distortion effect, perhaps one that might not be explainable using conventional theory, the kind that can lead a researcher to fame and his company to financial success. Of course, great effort and repetition of the test is necessary in that case to assure that such results are real and not due to limitations in the test setup or software limitations. A good way to verify the setup is to do a “dummy test” where two recordings made under identical conditions are compared. The resulting difference recording from such a dummy test can be used as a reference to compare with the result from the actual test.

3. PROGRAM DESIGN APPROACH

Audio DiffMaker provides for operation in single or dual channel mode with either 16 or 24 bits resolution. It can record, process, or play tracks at sample rates of 44.1kHz, 48kHz, 96kHz or 192kHz.

In the Audio DiffMaker software, some terminology has been coined to help organize and document program usage and operation. Up to three soundcards can be used in the system, these being the “Playback/Monitoring”, “Recording”, and “Source” sound card devices. They may however be (and usually are) all the same soundcard. Use of the “Source” device is optional, and is intended for providing stimulus audio for the test from the soundcard’s DAC. This is particularly helpful when it is the same soundcard (and locked at the same sample rate) as the Recording device. When that arrangement is usable, it provides automatic sample synchronization and avoids the need for sample rate compensation.

The two WAV recordings (“tracks”) that are recorded using the Audio DiffMaker system are designated as the “Reference” track and the “Compared” track. Equalization, gain adjustment, delay, and/or sample rate compensation may be applied by Audio DiffMaker to the Compared track for matching to the Reference track to the best of the program’s ability. The Reference track is then subtracted, sample by sample, from the processed Compared track to yield the "Difference" track. The Reference track remains unprocessed and is left intact. This avoids concerns that the program may be only comparing one processed track to another one, with questionable results. If there were a particular quality or "magic" in one of the tracks, and Audio DiffMaker had managed to match it to give a silent Difference result, then that quality must have merely have been one of the mundane processes that were applied by the program.

There are a number of "forms" (windows) that are available in the program, and only one will be visible at any given time. The first window that is available after startup is the "Settings" form where the user can choose soundcards, channels, rates, etc. There is an "advanced" option on the settings form which also provides access to more complicated settings that can select or control various parameters of track compensation.

The "Main Form" is configured to allow easy selection of WAV files, for use in difference extractions. It also provides a convenient scheme for playback and for comparing the sound of the various files. Each WAV file has an associated "Play" button that can be used even when another file is already playing. Rather than playing both together when that happens, the program will seamlessly switch playback to the newly selected file, at the same gain setting and at the equivalent time of the already-playing file. This allows the user to easily compare the files under equivalent conditions. This is particularly important when listening to Difference tracks so that a determination can be made about their audibility when played at the same gain levels as used to listen to the Reference track. There is also a provision to boost the playback gain by up to 70dB so low residuals and noise floors can be heard.

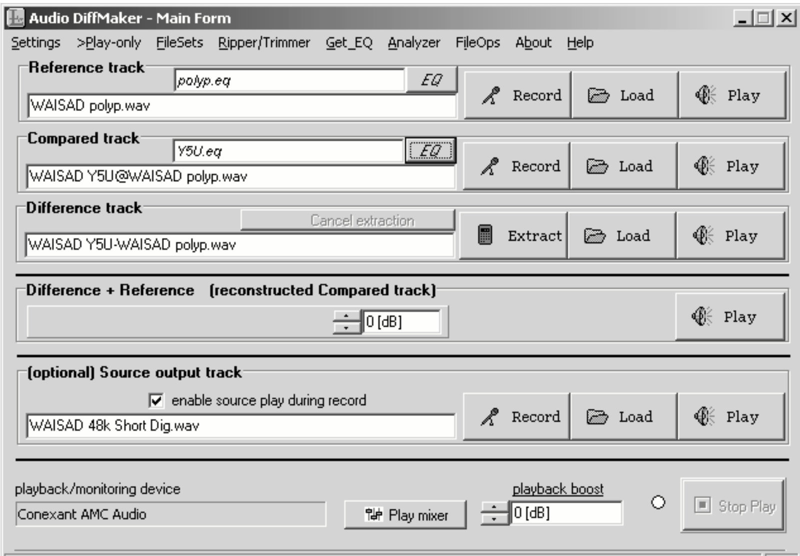

Figure 1: Audio DiffMaker's "Main Form"

Several other forms can be reached from the Main Form, including windows that provide user interfaces for making recordings, obtaining equalization (impulse response) characteristics or trimming WAV files. Also included is a convenient synchronous frequency response analyzer feature that can be used to inspect for even very small response variations of audio paths. For efficient sharing of associated WAV and EQ files for later comparison, a "File Sets" form provides a means for lossless compression and combination into (or later decompression of) special ".DYF" files useful for documenting experiments and their results.

4. PROGRAM OPERATION

A diagram of the recording setup is shown in figure 2. During use, the experimenter first chooses two conditions of an audio system path that he wishes to compare for possible effects on audio signals. If the difference between the paths is one that is expected or found to affect linear frequency response, the user should obtain a set of "EQ" files for both setup conditions. This is done by passing a swept sine wave signal through each path using a special program form to record and process the data. The swept sine wave stimulus used for measuring the EQ data can be provided from the soundcard or can be generated and transferred to other media in the form of a WAV file.

For each of the Reference and Compared tracks, a selected piece of program material (music or test signals), on the order of 10 to 60 seconds long, is recorded as it is played through each path. If the recordings cannot be made in a way that provides matched sample clocking of the stimulus hardware and the recording hardware, then these recordings should all be made within as close a time frame as possible to minimize effects of sample clock frequency drift.

After the various files are obtained, they are selected into the proper fields of the Main Form and the Difference track extraction is performed. This process can be time consuming, depending on computer speed, available memory, and file characteristics. During extraction, copies of pieces of the tracks are transformed back and forth many times between frequency and time domains for analysis and for iteratively finding the various parameters needed to best compensate the Compared track for uninteresting differences. A "Status" form that appears below the Main Form provides a running listing of the processes being performed so progress can be monitored.

When extraction is complete, a compensated Compared file replaces the original one in the Main Form and the Difference file is presented. The Reference file remains unchanged. The experimenter can listen to the set of files through the "Playback/Monitor" soundcard,

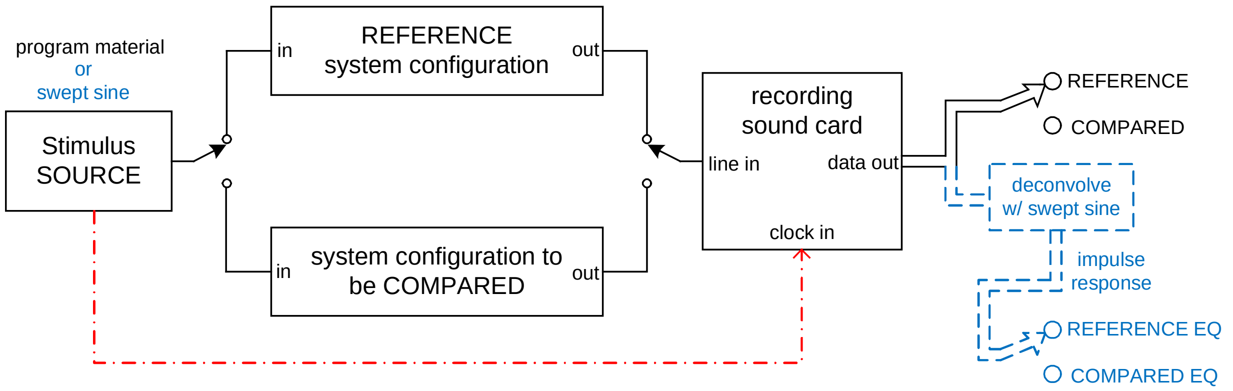

Figure 2 Connection diagram for acquiring the recordings

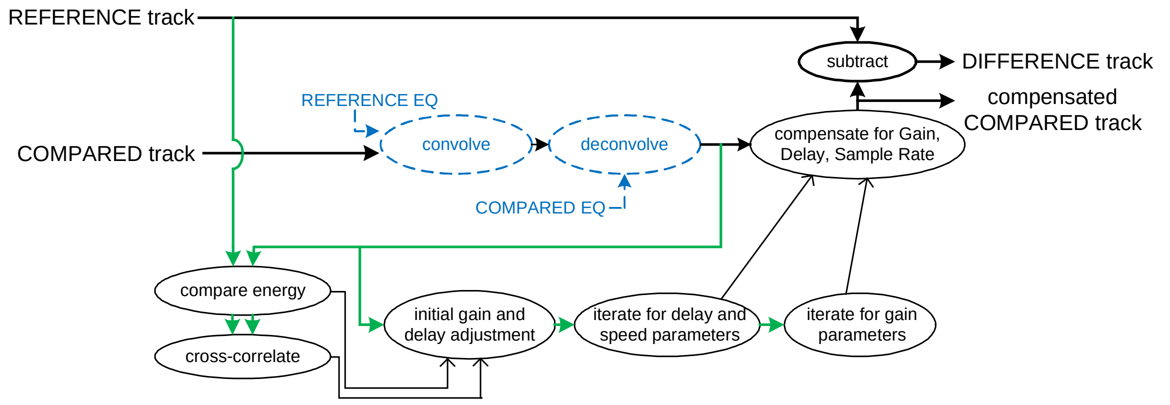

Figure 3 Compensation and difference processing

and if desired, save them for later combined as a set into a single ".DYF" file.

5. COMPENSATING FOR UNINTERESTING DIFFERENCES

A diagram of the processing steps is shown in Figure 3. Initial compensation parameters are first approximated using relatively simple methods. Level differences are estimated by comparing energies in the tracks, and delay offset is found within one sample along with polarity by performing a FFT-based cross-correlation between the tracks. These are used to make preliminary adjustments to copied portions of the Compared track. Gain and delay (and sample rate) parameter determination is performed iteratively, using a trial and error narrowing-down strategy. Each successive test tries a smaller deviation from the best-so-far parameter being optimized. Performance is driven toward the lowest absolute value of the zero-indexed cross-correlation between sample pieces of the Reference and trial Difference tracks. This was found to give superior results in the presence of noise than evaluations based on total energy in the Difference. Evaluations can be restricted to a selected frequency range to avoid degraded parameter determination due to noise in the less-audible frequency extremes.

The next subsections will provide brief descriptions of the requirements, processes and strategies used to compensate for known, uninteresting effects.

5.1. Frequency Response

Linear response compensation (equalization) is performed whenever related EQ files have been provided to the program for both the Reference and Difference tracks before extraction. If used, equalization is done only to the Compared track and before the other compensations. The EQ files contain records of impulse responses that were obtained by applying a periodically repeated log-swept sine wave (approximately 5 seconds long) to the tested setups, recording the outputs, and then deconvolving one period from each recording with the original swept sine wave. The obtained impulse responses are rotated within their record length so that the peak of the response occurs several milliseconds from the beginning of each record.

The impulse response related to the Reference track is first convolved with the Compared WAV data. Then that result is deconvolved with the impulse response from the original Compared track, thus imposing the linear response effects of the Reference condition while removing those that were only from the Compared condition. Care must be taken during deconvolution (a bin-by-bin division in the frequency domain) to avoid dividing by very small values, particularly at the frequency extremes, as this can produce noise in the result. Should that situation be detected, division using the offending sample is skipped.

A problem can arise if the effective sample rates during impulse response (EQ) acquisition were significantly different due to clock speed differences. This can degrade the usability of the equalization result, and a practical solution for that situation has not yet been devised.

5.2. Delay Compensation

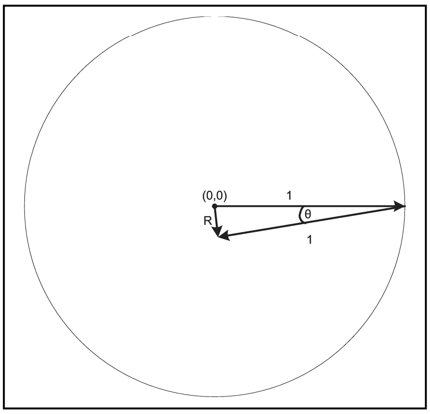

Figure 4 Diagram used to derive residual level from difference between two unit phasors

The sample points of the Reference and Compared tracks must be very carefully aligned before signal subtraction is performed. The needed amount of time position shifting is easily found and corrected to within one sample period at the midpoints of the tracks using conventional FFT-based linear cross-correlation. But that is far from being close enough to achieve significant null depths. This is particularly so at higher frequencies, where very small delay mismatches can equate to significant phase errors.

A non-zero complex spectral component of a recorded track at any given frequency can be represented by a phasor, normalized to unit length and assumed here to be at zero reference phase. Then subtraction of another vector of equal length but with a small phase error θ from the initial phasor will yield a resultant phasor as shown in figure 4. The length R of this resultant represents the normalized amplitude of the residual component left after the subtraction operation.

In phasor notation, the residual magnitude R can be expressed as

$$R = \left|1\angle 0 - 1\angle\theta\right|$$

In rectangular notation, it is

$$R = |1-(\cos\theta + j\sin\theta)| = \sqrt{(1-\cos\theta)^2 + \sin^2\theta}$$

$$R = \sqrt{2-2\cos\theta}$$

or in decibels relative to the original level:

$$10\log(2-2\cos\theta) \quad [dB]$$

This relationship is graphed in figure 5. To achieve a null depth of 50dB would require that the phase match be better than 0.2 degrees, a delay of only about 185 nsec at 3kHz. This is considerably less than the time between samples with any of the supported sampling rates.

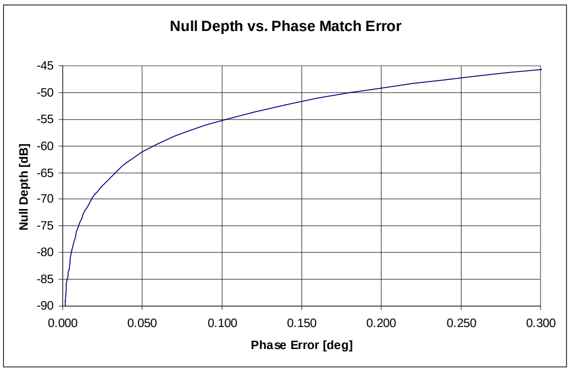

Figure 5 Limits to null depth with a given phase match error

In the time domain, a digitally recorded track can be easily shifted only by whole samples. But if it is transformed into the frequency domain, delays can be easily varied by any amount by adjusting the phase of each component an increment proportional to its frequency. For this reason, trial and error iteration to optimize delay compensation to fine values is done in the frequency domain.

5.3. Gain Compensation

Amplitude or gain differences can also substantially affect the residual level from a subtraction process. Gain variations can arise from the setups being tested as well as from voltage reference drift in digital converters.

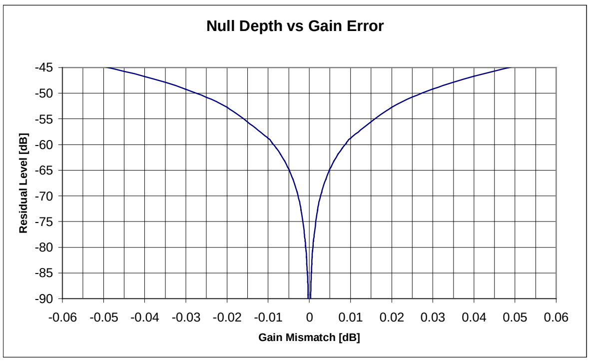

Figure 6 Limits to null depth with a given amplitude match error

If the degree of gain matching is E decibels, then the residual magnitude in decibels is

$$R = 20\log(|1-10^{E/20}|) \quad [dB]$$

The gain matching relationship is shown in figure 6. To achieve 50 decibels of null depth for any frequency component, the gain match must be within 0.02 decibels. For 70 decibels depth, better than 0.003 decibels of matching accuracy is required. Gain is easily varied by scaling in either the time or frequency domain. Fine optimization of gain parameters is done in Audio DiffMaker after the time alignment has first been completed and applied to a track section.

5.4. Sample Rate Compensation

Errors due to sample rate differences were not anticipated when work on Audio DiffMaker was begun. But early Difference tracks were noticed to often have substantially lower sound levels near the middle of each track, with higher levels of tinny sounding residuals near the beginning and the end. This effect would be worse when longer times had passed between when recordings were made or when recordings were longer in length.

A typical crystal oscillator as might be used for clocking in a soundcard or a CD player may drift on the order of 0.1 ppm over relatively short times. For a pair of 20 second Reference or Compared recordings in which the rates differed by 0.1ppm, the beginning or ending points would be about 10 seconds from a point midway through the tracks where simple delay alignment would typically be most successful. After 10 seconds, a 0.1 ppm clock speed mismatch would equate to 1 microsecond of misalignment, more than 1 degree of phase error in the sensitive 3kHz audio region, limiting the null depth there to only about 35 decibels.

The best way to avoid this kind of speed error is to arrange for the source of the audio test signal and the device used to record the Reference and Compared tracks to share a common sample clock as previously described. This is not always convenient or possible. To handle situations where clocks cannot be controlled like that, Audio DiffMaker is designed to analyze speed errors and implements a sample rate/position converter. The converter can be configured for various degrees of precision to be applied to the Compared track, trading off against processing time. Even at its default setting which begins by up-sampling the signal by 4 times the sample rate, this has been found to cause little degradation to achievable null depths. An example is shown in figure 7 of a Reference track and two Difference track results (note the vertical scale change of 30X) with and without sample rate compensation. In the center trace the phenomenon of poor nulling at the extremes of the signal is clearly visible.

The present implementation of sample rate compensation in Audio DiffMaker is intended to also address sample rates that smoothly drift during the times recordings of the Reference or Compared tracks are being made. Delay compensation parameters are found separately, using the previously described technique, for three different sections of the recordings (near beginning, near middle, and near end). The optimum time shift values found by this process, along with the sample index values associated with the centers of these three track sections are used to derive a second order mapping equation. The equation maps sample indexes for a new compensated Compared track to the effective indexes (usually non-integer) from which time matched data in the existing Compared track is to be obtained. The coefficients of the equation are then applied to generate the new re-sampled Compared track using precision up-sampled data and spline interpolation. This algorithm also inherently compensates for overall delay mismatches.

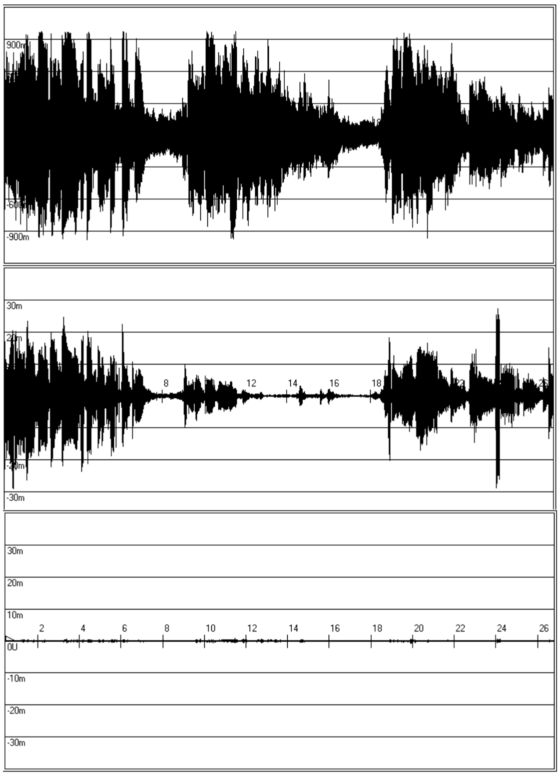

Figure 7 Effect of sample rate errors on residual pattern. Top: Reference track signal. Center: Residual with slight sample rate error. Bottom: Residual with sample rate compensated

6. APPLICABILITY

There are situations in which the digital difference extraction technique has not been successful. If there are high noise levels, these not only appear in the residual but can also limit the effectiveness of the program in finding proper compensation parameters for avoiding unwanted differences. Tests made to date using microphones recording output signals from loudspeakers have not been successful because of noise and perhaps other factors. Tests using analog signal sources such as tape recorders or phonograph turntables are unlikely to be repeatable enough to be successful. When any test results in a significantly audible Difference track, a "dummy" test should also be run to verify that the setup is even capable of producing two virtually identical recordings.

There have been some objections to the use of computer soundcards in this test. The concern is that the quality of the soundcard may be worse than that of the other components in the system, limiting the overall "resolving" power. This argument suggests that there is a limitation common to both the difference test and A/B listening tests should both signal paths involved have been changed too much in the same way. In comparative listening, the listener can be expected to be more sensitive to aspects of sound that seem relevant or familiar to him, such as whether a piano sound is realistic. Inclusion of lower quality components might limit the ability of a listener to detect subtle details because, to a human, one garbled sound can sound much like another. There is also the phenomenon of audio masking, in that a stronger sound effect may prevent a weaker one from being heard. Neither situation has much relevance to difference tests, however.

The difference test doesn't detect just aurally relevant changes, it detects audio band changes of *any* kind. The recorder need not respond to the sound with highest fidelity, it need only preserve an audible response to a change in sound. Only if an actual difference in the sound signal were something that is unrecordable could it be able to produce two identical recordings. Such a thing might happen should, for instance, a tested configuration produce high levels of radio frequency energy that would be removed by the bandlimiting digital audio recording process. The high frequency energy might degrade performance of a downstream component (such as a power amplifier) in an audible way. The solution for this is simply to record the signals after they have also passed through any affected component (the power amplifier in this example) that is suspected of being affected.

A similar approach can be taken when testing devices such as cables that may interact with other components. The signal can be recorded at a point following any components in the signal path that might be affected.

Through use of this software, differences due to passing an audio signal through different types of series capacitors have been isolated. Similarly, a lack of difference has been found (with little surprise) to result from painting the rim of a Compact Disc with a green felt-tip pen, a much-publicized audio "tweak".

7. CONCLUSION

A Windows based application has been developed and placed into the public domain, to allow recording and sophisticated comparison of WAV files for determining whether they might differ in audible ways. The method is intended to uncover whether differences are made to audio signals by treatments or devices. Types of differences that may be expected and not of interest were discussed as well as the algorithms used by the program to avoid them. It is hoped that use of this program will provide clearer evidence about whether effects reported to be audible actually result from changes to audio signals.

The Audio DiffMaker application may be freely downloaded from www.libinst.com.

8. REFERENCES

[1] Lipshitz, S. and Vanderkooy, J., "The Great Debate: Subjective Evaluation", J. Audio Eng. Soc., Vol. 29, No. 7/8, 1981July/August pp. 482-491.

[2] Baxandall, P., "Audible amplifier distortion is not a mystery", Wireless World, November 1977, pp.63-66.

[3] Hafler, D., "A Listening Test for Amplifier Distortion", HiFi News & Record Review, November 1986, pp. 25-29.

[4] Dunn, C. And Hawksford, M., "Toward a Definitive Analysis of Audio System Errors", presented at the AES 91st Convention, New York, USA, 1991 October 4 – 8.

[5] Farina, A., "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique", presented at the AES 108th Convention, Paris, France, 2002 February 19 - 22.