Time Varying Loudness as a Means of Quantifying Loudness in Nonlinearly Propagated Acoustical Signals

by

S. Hales Swift

Kent L. Gee, faculty advisor

A capstone project report submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Bachelor of Science

Department of Physics and Astronomy

Brigham Young University

April 2007

TIME VARYING LOUDNESS AS A MEANS OF QUANTIFYING LOUDNESS IN NONLINEARLY PROPAGATED ACOUSTICAL SIGNALS

S. Hales Swift

Department of Physics and Astronomy

Brigham Young University

ABSTRACT

Interest in predicting the community impact of flyover noise from high-performance aircraft has stimulated research into metrics that accurately represent human perception of the loudness of the noise thus created. This problem is complicated by both human factors involved in the perception of the noise waveforms, and physical ones, particularly the nonlinear process that underlies the transmission of high-amplitude noise through the atmosphere. An introduction will be given to the process of nonlinear propagation and how it is modeled to give our simulated nonlinear waveform. The processes underlying loudness perception in the hearing system and the methods used in this project to model them will be discussed. The reaction of the time-varying loudness metric, as described by Moore and Glasberg, to simulated linearly and nonlinearly propagated signals will be discussed and compared to its reaction to a "rephased" signal, in which the power spectrum of the nonlinearly propagated signal is maintained but the Fourier phase randomized to eliminate the shocks. Qualitative perceived reactions to the waveforms will be compared the quantitative reaction of the metric. These will be considered in the light of recent research into the phenomenon of "crackle" and the importance of the first derivative of the wave form in the qualitative experience of crackle.

Introduction

Background

Predicting the community impact of jet noise is a complicated problem involving both physical and human variables. This is a result of the interaction of both physical variables, such as atmospheric conditions and high sound levels, and physiological variable involved in the function of the human auditory system. Many metrics have been devised to deal with the impact of jet flyover noise including D-weighted loudness, Perceived Noise Level, and Stevens Mark Seven Loudness.

As understanding of the properties of the auditory systems has increased new metrics have emerged that propose to represent important features that were previously neglected such as temporal summation. An increased understanding of high-amplitude noise propagation and its unique effects has led us to try new methods of describing its impact on human beings. .It is to one of these, time-varying loudness, as well as its relationship to the associated problem of crackle that this paper is devoted.

Nonlinear Propagation

The physical complications come about as a result of the exceptionally high amplitude of jet noise waveforms. The nonlinear processes involved in propagating a high amplitude acoustical waveform through the atmosphere lead to steepened features in the waveform. [see figure 1] A high-amplitude signal experiences an amplitude-dependent speed of sound. As the amplitude increases the speed of sound increases. For the range of physical phenomenon dealt with in jet noise the first two terms of the Taylor series expansion are sufficient to qualitatively describe the results, thus the speed of sound may be given as

$$c = c_0 + \beta u$$

In this equation c0 is the small amplitude speed of sound 343 m/s, beta is a constant relating to the fluid, and u is the particle velocity. Acoustic pressure is proportion to particle velocity via Euler's equation in the far field. Thus pressure condensations advance faster than pressure rarefactions, resulting in waveform steepening, [see figure 1] and leading to the formation of acoustic shocks. These acoustic shocks affect the listener in much the same way as sonic booms, which are, similarly, also in the N-wave family.

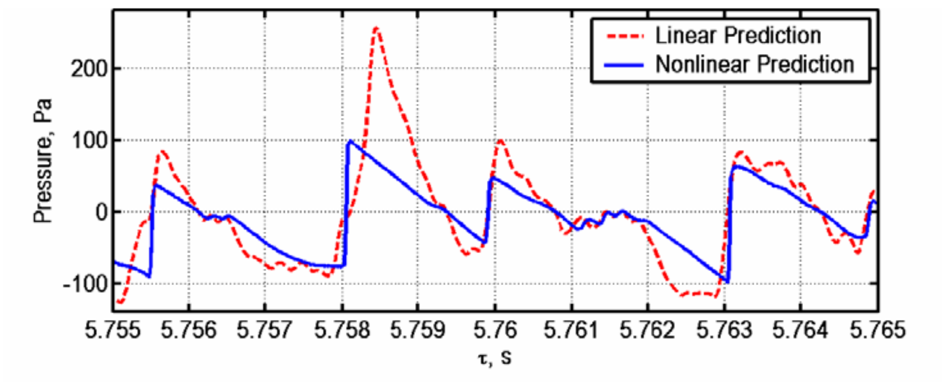

Figure 1. This is a snapshot of linearly and nonlinearly propagated waveform, note the steepening of the waveform in the nonlinear case and also the general attenuation of the amplitude.

Analogous to this is the case of a nonlinearly propagated sine wave, which evolves into a sawtooth wave in the absence of absorption. Due to the formation of the shocks and the attenuation of the rest of the wave, the loudness is concentrated in the time-domain into a number of discreet events. This is in contrast to the case of a steady-state noise waveform (one that sounds consistently the same qualitatively and has roughly the same amount of energy per unit of time, and average spectrum).

Previous Research into Metric Limitations as a Criteria for Determining the Usefulness of a Metric

Most measures of loudness or annoyance deal solely with the spectral content of the waveform. Previous research has proven this to be a critical limitation when dealing with nonlinearly propagated signals. This was shown by taking a nonlinearly and linearly propagated waveform, Fourier-transforming them and randomizing the phase of each spectrum in conjugate pairs to create a second real signal for each input with random phase (corresponding to a time-shift in each frequency component), but with the same power spectrum as the original signal. The rephased linear signal was then compared to the linear signal, and they were found to be qualitatively the same, and both fit the description of steady state signals. On the other hand, the rephased nonlinear signal and the nonlinear signal were found to be qualitatively quite different from one another. Thus nonrandom phase is seen to be an important characteristic of nonlinearly propagated signals and any metric that examines only the power spectrum will be unable to fully represent the unique perceptual impact of a nonlinearly propagated signal. As a result it is important in any metric that purports to capture the impact of a nonlinearly propagated signal to be able to address in a meaningful way the unique time-behavior of the acoustical waveform. [1]

“Crackle” as a Characteristic of Nonlinear Propagation and as a Possible Cause of the Qualitative Differences and the First Derivative as a Possible Explanation of the Differences Between the Nonlinearly and Rephased Nonlinear Waveforms

Another way of describing the unique quality of a nonlinearly propagated waveform is the phenomenon of “crackle”, defined as “sudden spasmodic bursts of a rasping fricative sound not dissimilar to that made by the irregular tearing of paper. . . . It is a startling staccato of cracks and bangs and its onomatope, ‘crackle,’ conveys a subjectively accurate impression.” [3] Recent work by Gee, et al suggests that crackle is associated with the first derivative of the pressure waveform rather than the actual pressure peaks of the waveform. [2] This then suggests that the qualitative difference between the nonlinearly propagated waveform and the rephased nonlinear waveform may be due to the removal of the acoustic shocks and the distribution of their energy to the rest of the waveform. It will thus be valuable to compare the values of the first derivative and the instantaneous loudness to determine if a correlation exists.

Complication in Modeling due to the Auditory System and Modeling Using Time-Varying and Stationary Loudness Methods

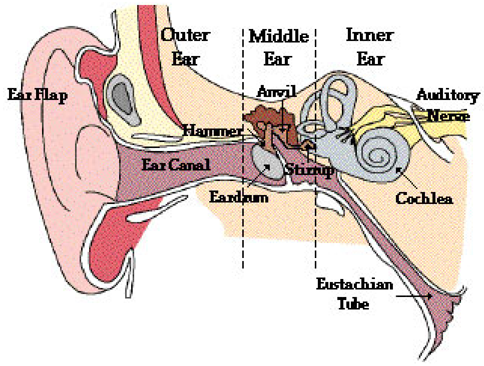

Figure 2 Diagram of the human auditory system

The human auditory system is, arguably, one of the most complex systems in the human body. [see figure 2] It converts energy from acoustical to mechanical, to fluid vibrational, to tortional energy, and transduces the resulting signal into nerve impulses before sending it up the auditory nerve to the brain for further summation and interpretation. Thus, it is not surprising that accurate modeling of human hearing is a process still undergoing revision and refinement. An accurate model of the hearing system begins with the transfer function of the head and outer ear. Some substantial differences occur in the spectrum entering the outer ear depending on what direction the sound comes from. These differences are due to the scattering behavior of sound waves around the head. Similar to the optical case, high frequencies experience less diffraction than low. The next step in the process involves transmission through the outer ear. The outer ear may be modeled as a closed-open tube of approximately 2.8 cm in length, with the usual frequency response (favoring odd harmonics). At the eardrum the pressure is converted into a mechanical excursion by the tympanic membrane. This motion is amplified mechanically in the ossicles and transmitted to the cochlea via the oval window. Each of these processes of amplification and transmission, along with the head related transfer function, has a frequency response associated with it. These are modeled collectively using a filter implemented in the frequency domain in both the stationary and time-varying versions of the loudness prediction program [see figure 3].

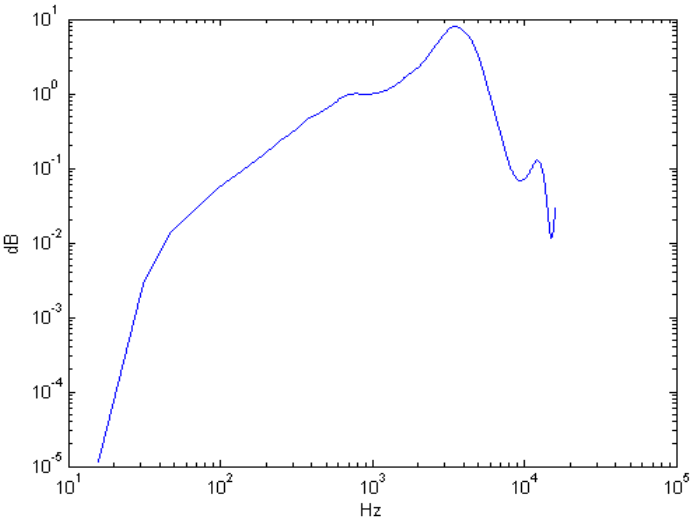

Figure 3 Magnitude of the filter response representing transmission from the environment to the cochlea for free field-frontal incidence

The next step in the process involves the basilar membrane and cochlea. Upon entering the cochlea (the sound is being transmitted via the cochlear fluid at this point), the vibration will excite different portions of the cochlea depending on what frequencies are present. Thus there is a frequency-place correspondence in the ear. The excitation of a particular region of the ear may be determined by modeling that region as a band-pass filter with center frequency the resonant frequency at that portion of the basilar membrane. This filter must be level dependent in order to accurately model the observed behavior of the cochlea [see figure 4, the response of level dependent filter centered at 1000 Hz]. This is implemented in MATLAB code by taking in the spectra and determining the power output of a bank of level-independent filters in order to determine the level reaching the cochlea in each Equivalent Rectangular Bandwidth (ERBn) of a normal hearing person. The equivalent rectangular bandwidth is essentially the bandwidth outside of which two signals no longer interact appreciably. The level reaching the cochlea is then used to determine the filter shape and the input spectrum is multiplied by the newly determined filters in order to come up with the cochlear excitation pattern. In order to obtain better frequency accuracy the filters and output are calculated at .2 ERBn intervals instead of 1 ERBn. The levels from each of the filters are then sent through a MATLAB function that represents the nonlinear growth of loudness in the ear, and accounts for the frequency dependent behavior of the cochlear amplifier. The output of this function is the specific loudness in each .2-ERBn interval and can be taken to represent the loudness due to the vibration of a particular length of the cochlea in response to a particular range of frequencies. The loudness in each $1/5^{th}$ of an ERBn is then summed to give the total loudness. In a steady-state sound, as described earlier, this may be taken to represent the total perceived loudness of the sound.

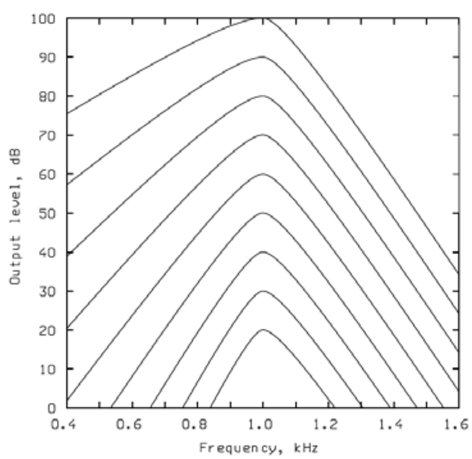

Figure 4 Response of the auditory filters centered at 1000 Hz to 1000 Hz sinusoids at levels from 20 to 100 dB in 10-dB increments, from ANSI S3.4-2005

Differences Due to Non-Stationary Signals

Up until this point what has been described is equally true of stationary loudness—as described in ANSI S3.4-2005 [4] and time-varying loudness [5]. However, many signals are not steady-state and the ear's temporal behavior thus becomes relevant. There are two notable effects that must be considered in order to effectively model the perception of a non-stationary signal. First, the loudness response takes time to build up [6]. For very short signals (neglecting some frequency smearing effects which may occur if proper ramps are not employed, i.e. time-domain tails that prevent the frequency band from becoming larger than a critical band), loudness grows towards the steady state as the length of time-waveform increases. Thus loudness summation is an important factor to account for in a model of loudness for non-stationary signals. Another important time-based behavior is temporal masking. This phenomenon is experienced when a loud sound is followed very shortly after by a quieter one. The quieter one is either not heard at all or is not perceived to be as loud in some cases as it otherwise would have seemed. These twin effects are modeled in time-varying loudness as described by Moore and Glasberg in their paper "A Model of Loudness Applicable to Time-Varying Sounds" [5].

Moore and Glasberg's Model of Time-Varying Loudness

In order to model time-varying loudness (TVL) as described in Moore and Glasberg's "A Model of Loudness Applicable to Time-Varying Sounds", one must begin with the time waveform instead of the spectrum. The time-series is first windowed in order to look at a very short portion of time. The length of time needed in order to examine different frequency ranges varies with the frequency of interest. For lower frequencies a longer period of time is needed to build up a meaningful loudness—corresponding, similarly, to the longer wavelength. On the other hand, a shorter window is needed to examine high frequencies because a high frequency component may be modulated many times in the time needed for a lower frequency to sufficiently build. TVL surmounts this obstacle by employing a bank of six Hanning windows of various lengths corresponding to the needs of the various frequency ranges. The outputs through these windows are then Fourier transformed to obtain spectral information for that moment. From these six FFT's, the running spectrum is obtained by taking the desired frequencies from each window. This running spectrum is then sent through the loudness summation process as described above (identical to ANSI S3.4-2005). This gives the instantaneous loudness.

The instantaneous loudness is believed by those responsible for TVL's creation to be a variable not accessible to human perception, but instead to represent a quantity similar to the total amount of activity in the auditory nerve at a given time. Thus further summation is necessary in order to represent the human impact of a waveform after being evaluated by higher cognitive stages. The next level up is termed short-term loudness and is believed to represent the loudness perceived by an individual at a particular time. The short term loudness (S') is calculated by taking the initial value S'(0) to be 0 and then comparing the current value of the instantaneous loudness (S) to the previous value of the short-term loudness. If the short-term loudness is less than the instantaneous loudness, corresponding to an attack, then the next element of the short-term loudness is found by

$$S'_n = (1-\alpha_{attack})S'_{n-1} + \alpha_{attack} * S_n$$

where alphaA is a time-constant, in this case corresponding to an attack. If S'(n-1)>S(n), corresponding to a decay then the next element of the short-term loudness is given by

$$S'_n = (1-\alpha_{decay})S'_{n-1} + \alpha_{decay} * S_n$$

In these equations, alphaD < alphaA, corresponding to the qualitative observation that we can hear an increase in loudness faster than we can hear a decrease.

Discussion of the Input Signals

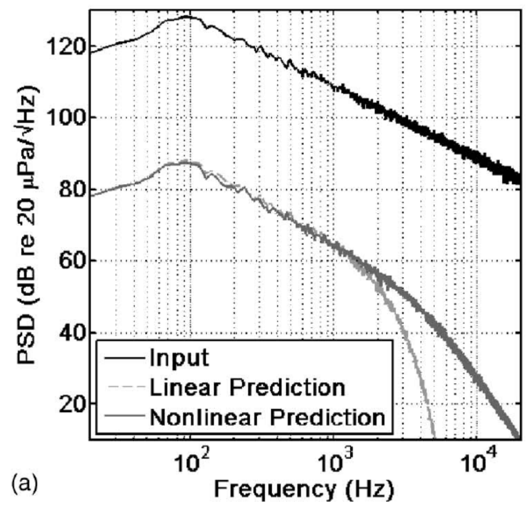

Figure 5 Frequency content of input signal and output signals, peak frequency of 100 Hz

It now seems appropriate to discuss the input signals and how they were generated. The initial input for all three of the signals investigated was a Gaussian random noise waveform. What is meant by Gaussian in this case is that its probability density function (the probability of the waveform assuming any particular value in its time series followed a Gaussian distribution). Likewise, the first derivative of the input time-series is also Gaussian—a truly random waveform with a white frequency distribution. This waveform is then filtered in the frequency domain to simulate the characteristically triangular shape of a jet mixing spectrum as viewed on a log-log plot [see figure 5]. The waveform is then obtained via inverse Fourier transform. It is then propagated using a linear model which takes into account atmospheric absorption—which has a low pass filtering effect as well as spherical spreading, but neglects nonlinear effects. It is simultaneously propagated numerically using a numerical method which solves the Generalized Burgers' Equation (GBE) and accounts for the previously mentioned. It does this by alternating between the time-domain, where nonlinear effects have their impact, and the frequency domain, where the atmospheric filtering is represented. In this way, both of these competing effects are modeled. This accounts for the creation of the first two signals used in this study. The third, the rephased signal was created by taking the nonlinear signal and Fourier transforming it and multiplying the positive and negative-frequency elements in conjugate pairs by a random (Gaussian) phase distribution and conjugate so that the inverse Fourier transform was real. This wave then had the same spectrum as the original nonlinear waveform (as was verified), and even a similar probability density function, but the wave no longer included the characteristic acoustic shocks. Listening to the waveform, there is a great qualitative difference between the two waveforms. The nonlinear one is dominated by the acoustic shocks, whereas the other is a steady-state signal sounding like a sharper sort of dryer noise.

Relevance of the Input Signals

The relevance of the first two signals is that they contrast the differences between the two modes of transmission linear and nonlinear. There are, however, many metrics that are able to make this distinction based on their spectrum again. The rephased waveform, since it has the same spectrum, but radically different temporal information, is an ideal candidate to weed out metrics that are not able to make the distinction between the signals that humans are able, easily, to make. This was also the signal used to show the limitation of spectrum based sound quality, loudness, and annoyance metric in the preliminary work, published in JASA-EL. This paper, along with the sound samples referenced in this report and that paper, are available for listening at the following link: http://scitation.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=JASMAN000121000001000EL1000001&idtype=cvips. The qualitative impact of these waveforms may be appreciated better by listening to them, probably, than in any other way.

Comment [a1]: I need to cite us here

Results

Short Term Loudness

The results contained both predictable features and a few surprises. The short-term loudness of the linear was less than that of the nonlinear, this expected based on previous results. The short-term loudness of the nonlinear signal was also less than that of the rephased nonlinear signal except at a few discrete moments [see figure 6]. As can be seen, the nonlinear signal has a larger amount of sudden variation in it than either the linear or the rephased signal. It was not initially expected that the nonlinear signal would have a lower short-term loudness than the rephased signal. As was mentioned, an increase in loudness is more easily observed than a decrease in loudness, so intuitively, we would expect to see the loudness of signal where most of the energy is concentrated in discrete boom events come out louder. However, it appears that the decrease in the noise floor, possibly combined with limitations in the model of loudness summation due to the extreme high levels that were attained in the instantaneous loudness (momentarily in excess of 115 phon), were sufficient to prevent the model from accurately representing the perceptual impact of the waveform—there is general agreement that the nonlinear waveform is either louder or in some other way more intense (possibly impulsive). In any event, although the short term loudness of the nonlinear signal occasionally surpasses that of the rephased signal, it is not generally the case that this happens.

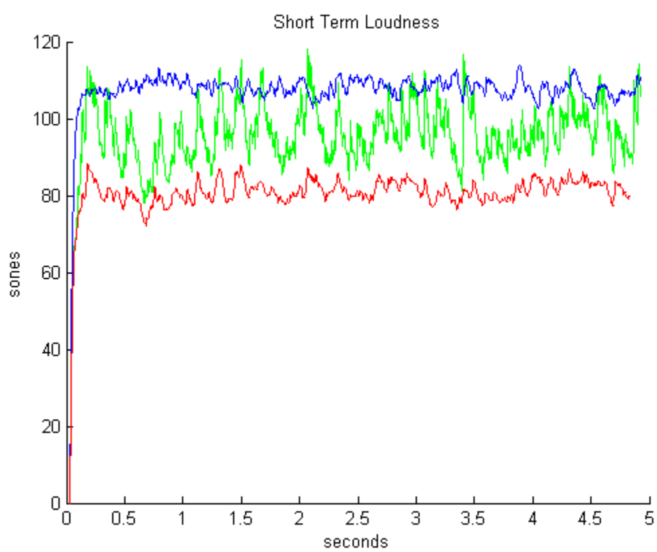

Figure 6 Short term loudness of the linear (red), nonlinear (green), and rephased (blue) signals

Instantaneous Loudness

As it turns out, the instantaneous loudness appears to be more representative of the perceived features of the nonlinearly propagated signal. The shocks are the most obvious feature to many listeners and this is not surprising given that their loudness momentarily approaches values of 115 phon. That this is the dominant *observed* feature is problematic, since, as was earlier stated, the instantaneous loudness is not believed to be a variable available for conscious perception, but is thought instead to represent the total activity in the auditory nerve at a particular instant. These results taken together with the conscious perception of the waveform suggest that either the instantaneous loudness may be available to conscious perception under some circumstances, or it may need a different scheme of summation if it is to be used for sounds containing discrete sound events.

Correlation with Crackle Work

Also the instantaneous loudness was strongly correlated with the first derivative of the waveform as can be seen in the figures [see figure 7 and figure 8]—even more so than with the peaks of the pressure wave form, indicating that the first derivative may be more influential in loudness perception for a highly impulsive waveform than the pressure peaks.

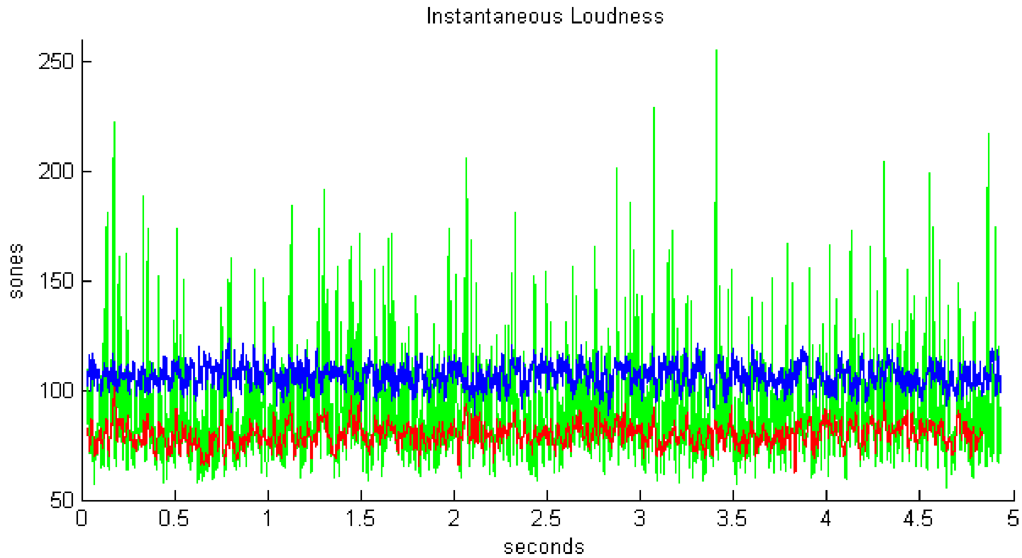

Figure 7 Instantaneous loudness of nonlinear (green), rephased (blue), and linear (red)

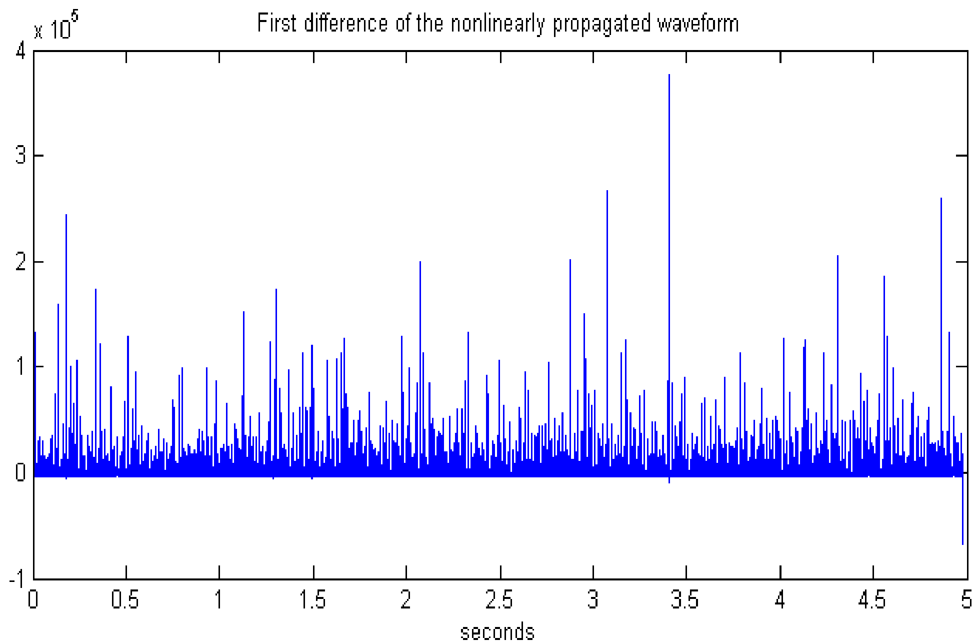

Figure 8 First difference of the nonlinearly propagated waveform

This is especially relevant in light of recent research on the "crackle" phenomenon. Recent research by Gee, et al. has found that the probability density function (PDF) of the first derivative of a crackling signal is highly skewed and has a long tail proceeding far into the positive values. This is of interest because it supports the observation that the distinctive events heard in a nonlinearly propagated signal are marked by acoustic shocks (these are what compose the positive tail of the PDF). Accordingly the PDF's of the first difference of the linearly and nonlinearly propagated waves has been taken [see figure 9]. The distinctive skewness and tail did indeed occur indicating that the first difference, whether examined as it occurs in the time domain or by way of its statistics, represents a distinctive differentiating feature between the nonlinear and the rephased waveforms. This unique feature appears to be the strongest indicator of crackle discovered to date.

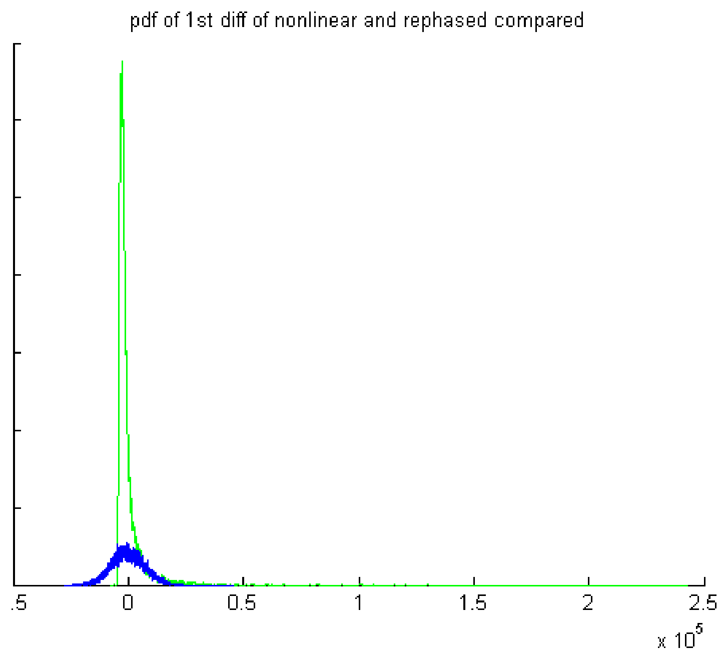

Figure 9 Probability Density Functions of the first differences of the nonlinear and rephased nonlinear signals (notice especially the long green tail which corresponds to the large spikes in figure 8)

Conclusions and Future Directions

As has been mentioned, the major loudness events of a crackling signal appear to be correlated with the extreme values of the first derivative. An entire waveform may be characterized in this way by taking the probability density function of the first difference. It would be valuable to examine the relationship between values in the first difference sequence and the time-varying loudness to attempt to determine an empirical functional relationship between the two. If this could be accomplished the time taken to calculate time-varying loudness (currently 6-8 hours for the five second waveforms used) could be greatly reduced. However, it is not known at this time whether this relationship holds for a broader class of signals or if it is unique to the class of crackling or highly impulsive signals. The apparent deficiency of short-term loudness in capturing the extraordinarily high peaks may suggest that the model requires further adjustment of the attack times in summation equations to deal with what may be aspects of temporal summation unique to this unusual class of signals. This could be by way of suggesting several attack constants of different magnitudes corresponding to the magnitude of the difference between the short term loudness and the instantaneous loudness, similar to what is currently done to determine whether an attack or decay is occurring—using inequalities. Thus time-varying loudness is a step in an illuminating step in the right direction but still requires further development and refinement if it is to be applied successfully to highly impulsive or crackling signals.

References

[1] “On the potential limitations of conventional sound quality metrics in quantifying perception of nonlinearly propagated noise”, Kent L. Gee, S. Hales Swift, Victor W. Sparrow, Kenneth J. Plotkin, and J. Micah Downing, J. Acoust. Soc. Am. 121, EL1 (2007)

[2] “On the Perception of Crackle in High-Amplitude Jet Noise”, Kent L. Gee, Victor W. Sparrow, Anthony Atchley, and Thomas B. Gabrielson AIAA JOURNAL Vol. 45, No. 3, March 2007

[3] “‘Crackle’: An Annoying Component of Jet Noise,” Ffowcs Williams, J. E., Simson, J., and Virchis, V. J., Journal of Fluid Mechanics, Vol. 71, No. 2, 1975, pp. 251–271.

[4] ANSI S3.4-2005

[5] “A Model of Loudness Applicable to Time-Varying Sounds”, Brian R. Glasberg, and Brian C. J. Moore J. Audio. Eng. Soc., Vol. 50, No.5, May 2002

[6] “Temporal Integration of Loudness, Loudness Discrimination, and the Form of the Loudness Function”, S. Buus, M. Florentine, and T. Poulsen, J. Acoust. Soc. Am. 101, February 1997