SEPTEMBER 30 2024

Reanalysis of crackle perception data using logistic and logarithmic fits and sound quality metrics FREE

S. Hales Swift; Kent L. Gee

Proc. Mtgs. Acoust. 50, 040009 (2022)
https://doi.org/10.1121/2.0001957

POMA Proceedings of Meetings on Acoustics

Volume 50 http://acousticalsociety.org/

183rd Meeting of the Acoustical Society of America

Nashville, Tennessee

5-9 December 2022

Noise: Paper 4aNS6

Reanalysis of crackle perception data using logistic and logarithmic fits and sound quality metrics

S. Hales Swift
Department of Photonic and Phononic Microsystems, Sandia National Laboratories, Albuquerque, NM, 87114; hales.swift@gmail.com

Kent L. Gee
Department of Physics and Astronomy, Brigham Young University, Provo, UT, 84602; kentgee@byu.edu

Prior work has shown a robust relationship between crackle from high-performance jets and the skewness of the distribution of the first time derivative of the pressure waveform; however, prior efforts have characterized this relationship in terms of a linear relationship between the log of the derivative skewness and the category scaling responses. While the relationship is linear over important portions of its range, the use of a logistic curve fit more fully captures the characteristics of the relationship between log first time derivative skewness and category scaling relationship including implied non-negativity, and saturation. Additionally, time-varying sound quality metrics including loudness and sharpness have shown sensitivity to jet crackle. Accordingly, the subject data of a prior study –in which the "crackliness" of jet sounds was rated– are re-analyzed using the derivative skewness of the pressure waveform and sound quality metric-based variables as potential predictors of a crackling sound quality. After accounting for nonlinear trends in some of the predictor variables, a relationship using the derivative skewness and the standard deviation of the sharpness is able to account for all but 1.1% of the variance in the mean subjective crackle ratings for the waveforms in the prior study.

Published by the Acoustical Society of America

https://doi.org/10.1121/2.0001957.

1. INTRODUCTION

A. BACKGROUND ON CRACKLE

Crackle describes the distinctly abrasive quality of high-performance jets;[1] many experience crackle at sporting events featuring flyovers by high-power tactical fighter aircraft without necessarily being aware of the technical term describing the phenomenon responsible for the subjective experience.[2] Crackle has long been of interest within the jet noise community because this sound quality is a significant contributor to the annoyance elicited by these sounds.[1] First described by Ffowcs-Williams,[1] efforts to assess it largely fell into two main approaches: measuring the skewness of the pressure waveform, $\mathrm{Sk}\{p\}$, and measuring the skewness of the derivative of the pressure waveform (hereafter, derivative skewness, or $\mathrm{Sk}\{\partial p/\partial t\}$).

Skewness is a measure of asymmetry in a distribution of values. Crackling waveforms contain a multitude of shocks—extremely rapid, almost instantaneous increases in pressure—with intervening periods of slower, smoother decreases in pressure. Consequently, the distribution of derivatives of a crackling waveform contains a majority population of modestly negative values corresponding to the gradual pressure decreases, and a smaller number of extremely positive values corresponding to the shocks. Because of the sensitivity of the measure to shock content, crackling jet noise waveforms exhibit elevated values of derivative skewness.[3] Jet noise waveforms also often exhibit positive pressure skewness,[1] though this does not seem to directly affect the sound quality.[4] The key developments in the effort to understand how to correctly measure crackle are summarized briefly in Reference 5.

B. GEE *ET AL.* STUDY INTRODUCTION

Arguments over the best physical measure of crackle culminated in a formal jury study on jet crackle perception, in which the two authors of this presentation were involved.[5,6] 31 listeners evaluated 15 sounds from the F-35, rating the degree of "crackliness" present in each 3-second sample using a slider. Replays and random access were permitted, and listener responses showed conclusively that the jet noise crackle percept is strongly associated with the derivative skewness, $\mathrm{Sk}\{\partial p/\partial t\}$, and that the skewness of the pressure waveform, $\mathrm{Sk}\{p\}$, falls into insignificance when they are analyzed jointly.[5] The analysis in that paper used linear regression with the logarithm of the derivative skewness to fit the data. It showed conclusively the efficacy of $\mathrm{Sk}\{\partial p/\partial t\}$ as a measure of the physical phenomena that lead to the crackle percept. The linear function fit presented in that work explained 93.3% of the variance. However, the choice of a pure linear function fit (including to the logarithm of the quantity) is questionable on several grounds, and can be improved upon as we will demonstrate in this work.

C. LOGISTIC FUNCTION INTRODUCTION

While a linear function fits their study data well, it is worthwhile to consider what happens outside of the range of derivative skewness values that were evaluated in the subject test, a limitation noted by Gee *et al.*[5] For example, because it is symmetrically distributed, a period of a cosine wave has $\mathrm{Sk}\{\partial p/\partial t\} = 0$. The logarithm is undefined for zero inputs, and complex-valued for inputs less than zero, resulting in a large class of signals—those with $\mathrm{Sk}\{\partial p/\partial t\} \leq 0$—for which the expression given in the original study is inadequate. This class of signals also includes any time series with a finite but sufficiently small derivative skewness, for which the linear relation with $\log_{10}(\mathrm{Sk}\{\partial p/\partial t\})$ results in negative values outside of the scope of possible responses in the subject test. Similarly, for large enough values of $\log_{10}(\mathrm{Sk}\{\partial p/\partial t\})$, the relation predicts outcomes greater than 50, the highest value possible in the subject test. The logistic function, on the other hand, overcomes these problems: its parameters can be chosen to yield a zero asymptote in the negative direction and an asymptote of 50 in the positive direction, consistent with the requirements of the experiment. The equation for the logistic curve is

$$r(x) = \frac{l_{upper} - l_{lower}}{1 + e^{-k(x-x_0)}} + l_{lower} = \frac{50}{1 + e^{-k(x-x_0)}}, \tag{1}$$

where $r(x)$ is the predicted rating, l_{lower} is the lower rating limit, 0, l_{upper} is the upper rating limit, 50, x_0 is the middle of the logistic curve, and $k\frac{l_{upper}-l_{lower}}{4}$ is the slope at the midpoint. These latter two parameters are conceptually akin to the intercept and slope in a linear equation. We thus end up with an equation which, similar to the prior linear fit, has two undetermined parameters to be determined using the data. However, the logistic equation's structure conveniently avoids the "out-of-bounds" issues of a linear function fit. This paper consequently takes as one of its priorities evaluating the improvement in fit that results from using a logistic function fit. In doing so, however, it is important to realize that the original work was not seeking to predict precise crackle ratings so much as to define boundaries between categories of crackle ratings; because of this, there was less rationale for determining the optimal function type in that effort, allowing that task to be taken up here.

2. REANALYSIS

A. REFITTING USING THE LOGISTIC FUNCTION

The data were refit using a logistic functional form with the lower and upper asymptotes 0 and 50, respectively, as dictated by the experimental constraints, and the linear and logistical function fits plotted in Fig. 1. The values of x_0 and k were chosen to minimize the RMS error; the RMS error is equivalent to the standard deviation of the residuals, so the typical difference between the predicted crackle rating and the average subject rating decreases by 24% when a logistic function is used instead of a linear fit. Thus, instead of the crackle rating prediction typically differing by 3.3 from the subject rated value, with the logistic fit it typically differs by 2.5. The logistic function fit avoids the prediction of a negative value for the first experimental data point, which the linear fit implies if used for predictive purposes. In the plot, the regions where the linear fit yields out-of-range values are shaded gray. Several groups of points in the figure have been marked with ovals; the logistic curve passes notably closer to the points in the first two groups that are marked by light blue (cyan) ovals. However, the group of points marked by the dotted red oval paints a more complicated picture because they are non-monotonic, potentially suggesting the presence of factors in crackle perception not fully accounted for by derivative skewness. In one instance, they are even non-monotonic with widely non-overlapping standard error intervals. Non-monotonicity is surprising because if we have:

1. a metric which completely describes human response to crackle, and
2. have faithfully reproduced the sounds based on that metric, then we would expect that

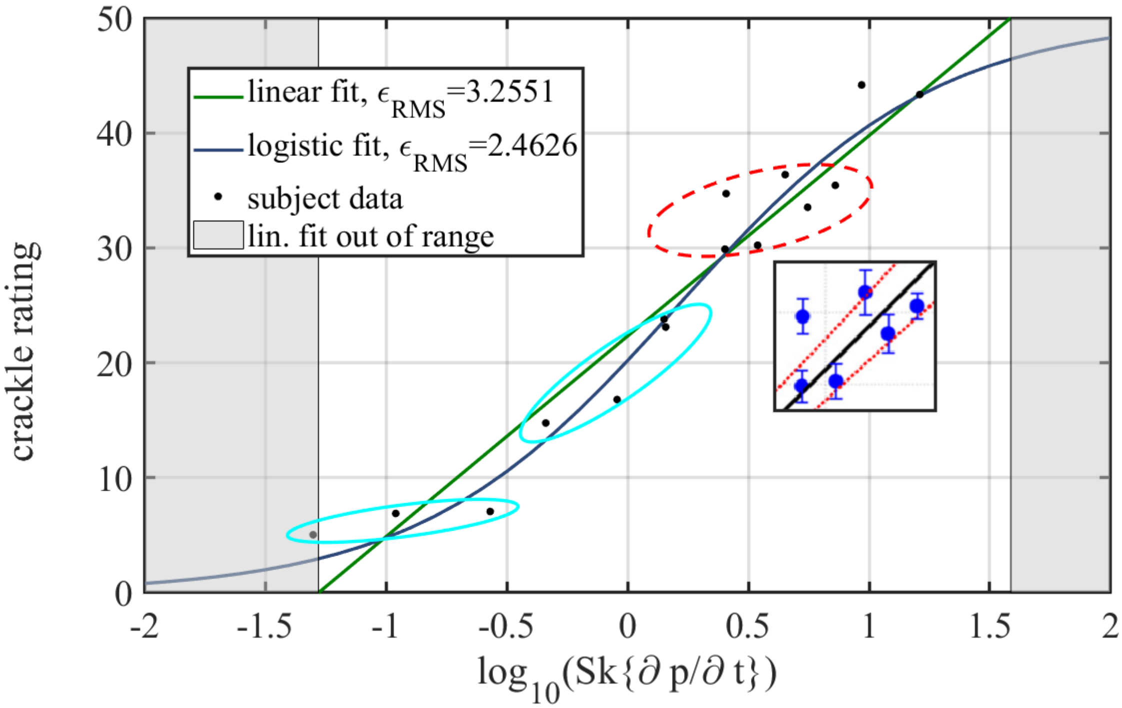

Figure 1: Linear and logistic fits to the human subject crackle ratings with respect to the base ten logarithm of the derivative skewness. The domain for which the linear fit gives of range values is shown in grey. The logistic fit results in a 24% decrease in RMS error.

3. a monotonic increase in the independent variable leads to a monotonic increase in the dependent variable.

A completely descriptive metric should preserve ordering under these circumstances. Human response is intrinsically noisy or uncertain, so some finite scatter in the data is expected. However, considering the second and third points in the inset of Figure 1 the substantially disjoint calculated error bars complicate an appeal to experimental uncertainty as an explanation.

B. REANALYSIS USING SOUND QUALITY METRICS

The above analysis, however, takes certain assumptions for granted. To begin with, it is worth asking what aspect(s) of crackle the derivative skewness, $\mathrm{Sk}\{\partial p/\partial t\}$, is measuring. It is a nearly ideal measure for assessing the presence of nonlinear steepening and shocks and, given adequate measurement system capabilities, assesses the presence of these features in the waveforms well. However, this approach risks focusing entirely on the physical processes which lead to crackle rather than on the attributes that make up the perceptual experience of crackle. The human hearing system has differing sensitivity to different frequencies of sound and, consequently, the same sound (with the same derivative skewness) that produced a crackling sound quality might be entirely inaudible if played back at too high of a frequency or undergo other more subtle alterations in perceived character under lesser shifts in frequency. Additionally, it is not automatically clear which aspects of the sound lead to the human experience of a succession of shock waves. Is the perceptual essence of a crackling waveform to be found in its temporal concentration of loudness, or the temporal concentration of high frequencies, or the coincidence of the two, or some other factor? Consequently, it makes sense to explore the possibility that the use of sound quality metrics—as mathematical models of key aspects of the hearing system—could enable an improved picture of crackle as a perceptual entity, with associated improved predictive capability, and all of the attendant community noise management benefits.

With these objectives in mind, we turn to a collection of metrics that have been shown to be associated with crackle in a prior informal listening study[7] that focused on the sound quality of altered jet noise waveforms; the metrics included in this prior work were:

- instantaneous loudness distribution,
- instantaneous sharpness distribution,
- distribution of the product of instantaneous loudness and instantaneous sharpness,
- correlation between instantaneous loudness and instantaneous sharpness,
- roughness, and
- sharp roughness.

Though extremely promising, the latter two metrics are not explored on this occasion because these codes, which implemented roughness based on the Cambridge loudness metric, have not yet been concretely validated by comparison with either human subject data or a metric well anchored to such data. Based on the results of the prior informal test, the following more adequately validated metrics were explored (with the theoretical rationale for each in parentheses):

- mean sharpness (control on spectral effects)
- standard deviation of instantaneous loudness (exploits loudness concentration),
- skewness of instantaneous loudness (exploits loudness concentration),
- skewness of short-term loudness (exploits loudness concentration),
- skewness of instantaneous loudness normalized by short-term loudness (exploits loudness concentration),
- standard deviation of instantaneous loudness effectively normalized by short-term loudness because the signals were normalized in loudness (exploits loudness concentration),
- standard deviation of instantaneous sharpness (exploits spectral concentration),
- skewness of instantaneous sharpness (exploits spectral concentration),
- skewness of instantaneous sharpness normalized by short-term loudness (exploits spectral concentration),
- standard deviation of instantaneous sharpness normalized by short-term loudness (exploits spectral concentration),

	Sk{S}	Sk{N}	Sk{S'}	Sk{NS}	σ_N	ρ_{NS}	σ_S	μ_S	σ_S/μ_S
Sk{S}	1.0000	0.8434	0.8608	0.7653	0.5437	0.7172	0.3160	-0.0721	0.3855
Sk{N}	0.8434	1.0000	0.9881	0.8278	0.6800	0.6240	0.3814	-0.4310	0.5186
Sk{S'}	0.8608	0.9881	1.0000	0.8590	0.7103	0.6568	0.4335	-0.3910	0.5612
Sk{NS}	0.7653	0.8278	0.8590	1.0000	0.8952	0.8901	0.8009	-0.0043	0.8518
σ_N	0.5437	0.6800	0.7103	0.8952	1.0000	0.7189	0.8830	-0.0679	0.9585
ρ_{NS}	0.7172	0.6240	0.6568	0.8901	0.7189	1.0000	0.7739	0.3723	0.7393
σ_S	0.3160	0.3814	0.4335	0.8009	0.8830	0.7739	1.0000	0.3017	0.9739
μ_S	-0.0721	-0.4310	-0.3910	-0.0043	-0.0679	0.3723	0.3017	1.0000	0.0933
σ_S/μ_S	0.3855	0.5186	0.5612	0.8518	0.9585	0.7393	0.9739	0.0933	1.0000

Table 1: Correlations between sound quality metrics.

- skewness of the product of instantaneous loudness and instantaneous sharpness (exploits coincidence between spectral and loudness concentration), and
- correlation between instantaneous loudness and instantaneous sharpness (exploits coincidence between spectral and loudness concentration).

To provide some context for these metrics, while not originally thought to be available for conscious perception,[9] the instantaneous loudness has been seen as useful in characterizing certain perceptual "textures" that may or may not be appropriately called loudness.[10] The short term loudness on the other hand is thought to be the loudness that one typically experiences at a moment in time.[9] In dividing the instantaneous loudness by the short term loudness at the same point in time, something akin to a coefficient of variation (the standard deviation divided by the mean of a quantity) is produced; however, unlike a more globally oriented coefficient of variation, the fluctuation in the instantaneous loudness is normalized by the local loudness (which varies significantly in many of the waveforms considered) rather than a less relevant global value. The standard deviation by itself gives an estimate of how much a quantity varies; for crackle, this is important because the shocks produce peaks in loudness and sharpness, while the spaces between the shocks result in lower values for both quantities than would be present in a noise spectrum with the same spectral content making the variation an indicator or crackle. The correlation measure uses the coupling between peaks in loudness and sharpness in crackling waveforms as a crackle indicator. Given the role of derivative skewness in crackle research generally and the prior evidence of crackling waveforms producing skewed distributions of these variables, these skewness-based measures and other metrics seemed appropriate.[7] These metrics can be calculated using codes found on the MATLAB file exchange.[8] The correlation values between the sound quality metrics considered are given in Table 1.

C. SOUND QUALITY METRIC PROCEDURES

In the original subject test, the loudness of the sounds fell within a specified range of 23.4 ± 0.6 sones. In order to ensure the loudness of the sounds fell within this range, the level of the sound was iteratively varied with the calibration factor $C_{23.4}$ updated as $C_{23.4} \rightarrow C_{23.4} \times (23.4/\mu_{loud})^2$

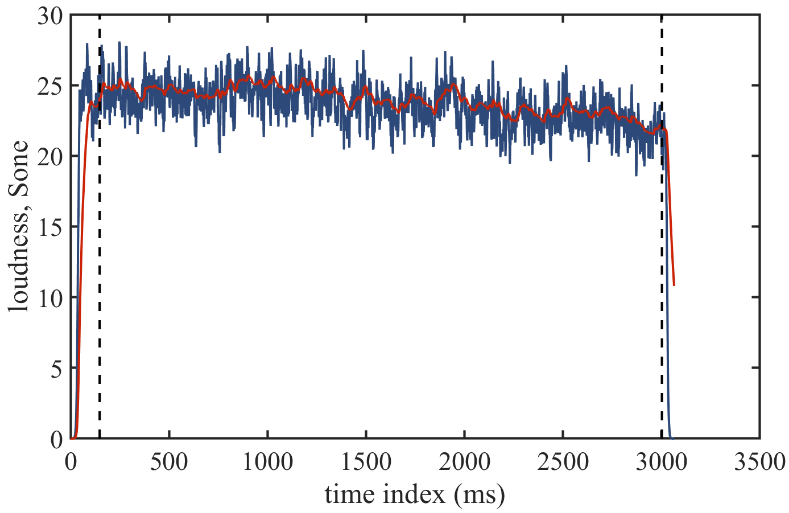

Figure 2: Example loudness trace with truncation limits shown. Outside the limits noticeable transient behavior associated with signal onset/offset occur.

where μ_{loud} is the mean *short-term* loudness for each signal. Beginning with a calibration value of $C_{23.4} = 1$ for all signals and following this process for each signal, the mean loudness of all signals was within 7.1×10^{-4} of the target value within two iterations.

The short-term loudness of a steady sound takes a certain amount of time to achieve its asymptotic value and similarly takes time to decay after signal cessation. In order to avoid the resultant transient periods associated with the beginning and ending of the signals, the first 145 and the last 62 samples of the metric variables were discarded from all analyses after identifying these transient intervals through visual inspection of all time series. The truncation bounds are shown for reference with an example loudness trace in Fig. 2. All subsequent analyses thus involve only the remaining intermediate values.

3. RESULTS

The metrics identified above were evaluated and their statistical associations with the dependent variable (crackle predictions reported during human subject testing) were assessed using linear regression. The relationship between derivative skewness and crackle prediction was also revisited in order to avoid introducing an unfair procedure by hunting for better curve fits using only the sound quality metrics, and ignoring derivative skewness in some of the more sophisticated analyses.

A. RAW SINGLE-METRIC PREDICTIONS

The single-metric predictors outlined above were calculated and processed as explained and the resultant relationships with rated crackle are shown in Fig. 3; in each instance where they occur, S is sharpness, N is loudness, μ is mean, σ is standard deviation, ρ is correlation, and Sk is skewness; primes on a variable indicate short-term rather than instantaneous values. As a starting point for evaluating which model variables proved most indicative in terms of explained variance, r^2, the subparts of the figure are given in ascending order of variance explained by a linear dependence on the proposed variable (i.e., ascending r^2 values); they are shown in Table 2 in the same order.

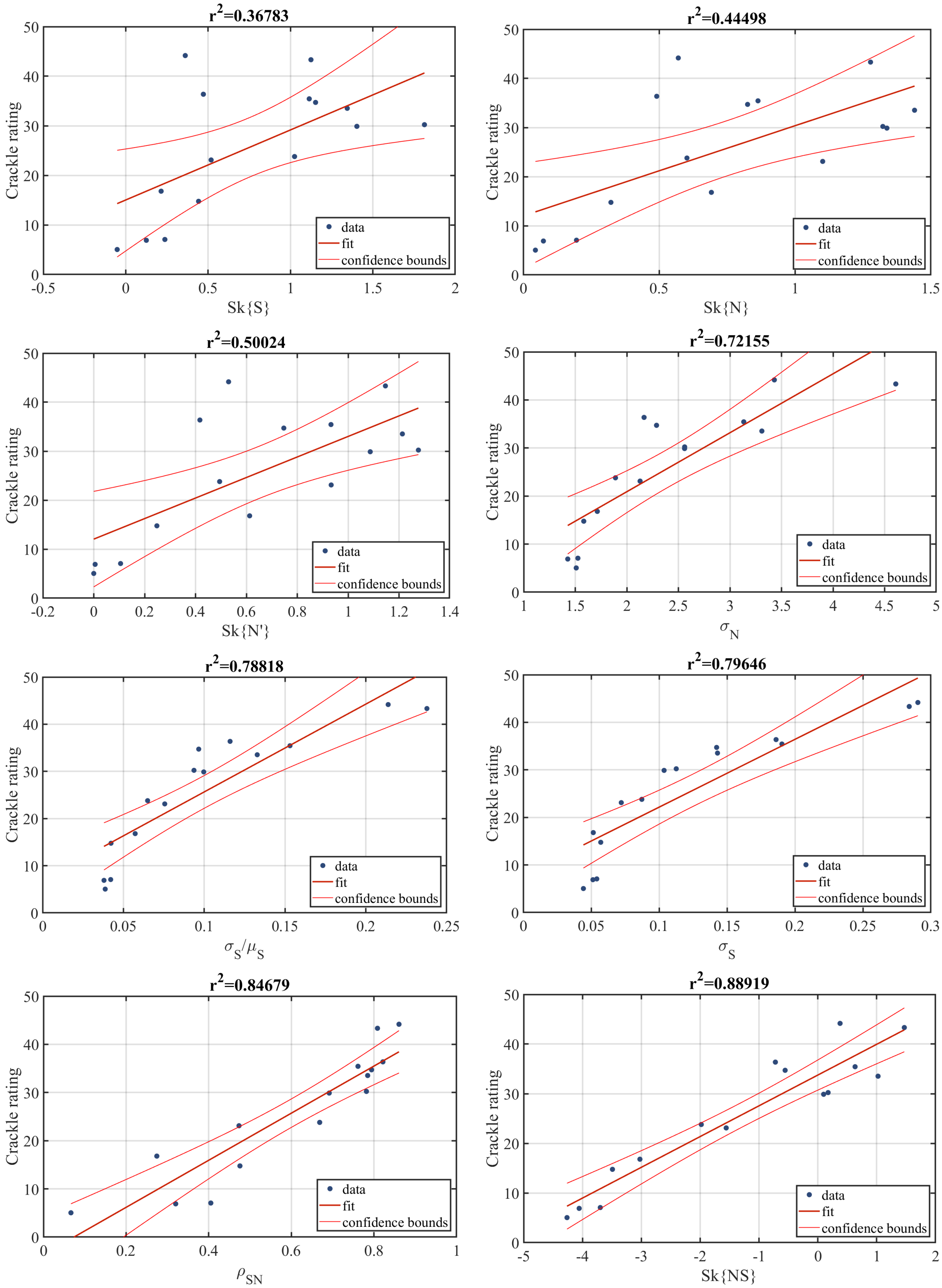

Figure 3: Raw subject crackle ratings shown as a function of each of the proposed sound quality-based predictors with linear fit and 95% confidence intervals. Variables are presented in order of increasing r^2 values. Nonlinear trends in data limit linear explanatory efficacy in some instances, particularly for σ_N and σ_S.

Metric	μ_S	Sk{S}	Sk{N}	Sk{N'}	σ_N	σ_S/μ_S	σ_S	$\rho_{S,N}$	$Sk\{NS\}$
r^2: linear	3.6%	36.8%	44.5%	50.0%	72.2%	78.8%	79.6%	84.7%	88.9%
r^2: $\log a + x$	N.C.	N.C.	N.C.	N.C.	90.0%	N.C.	95.7%	N.C.	90.8%

Table 2: Variance explained by linear and $\log(a + x)$-form nonlinear single-variable sound quality-based models. N.C. indicates quantity not calculated because nonlinearity did not appear to be a main issue.

The mean of sharpness was not expected to be particularly indicative of a crackling sound quality because the variation rather than the average quantity is believed to be the important variable, but seemed worth testing in case it affected the relationship in some way. However, it accounted for only 3.6% of the variance and so is not shown. $\mathrm{Sk}\{S\}$ and $\mathrm{Sk}\{N\}$ are modest predictors of crackle, predicting 36.8% and 44.5% of the variance, respectively; however, their product is a strong predictor of crackle, predicting 88.9% of the variance. A likely explanation for this is that the shocks in crackle-containing waveforms lead to concentration of both sharpness and loudness in time, so the skewness of their product, $\mathrm{Sk}\{NS\}$, exhibits much greater sensitivity to crackling sound quality than either variable independently, predicting 88.9% of the variance. The correlation coefficient between the instantaneous loudness and the instantaneous sharpness, ρ_{SN}, similarly performs well by exploiting the strong relationship between loudness and sharpness in crackling signals, predicting 84.7% of the variance. The concept of a coefficient of variation (the standard deviation of a quantity divided by its mean) seemed potentially useful in this venue. Consequently, the relationship of the sharpness coefficient of variation to the response variable seemed worth including, and accounted for 78.8% of the variance.

The standard deviation of loudness, σ_N, and sharpness, σ_N, were both fairly strong predictors of crackle, predicting 72.2% and 79.6% of the variance, respectively, and also make intuitive sense: crackle results in large variations in both loudness and sharpness. The standard deviation of loudness is also a de facto coefficient of variation because the loudness of sounds was kept uniform in this test and thus gives us a qualitative idea of the *proportional* impact of this variation on the texture of the sound. Despite the relatively large portion of the variance explained by linear models on σ_N and σ_S, it is immediate clear that the relationships between these metrics and crackle perception are quite nonlinear and, consequently, the r^2 value may not truly capture the strength of a potential causal relationship without further processing.

B. LOG-TRANSFORMED SINGLE-VARIABLE PREDICTORS

This caveat regarding nonlinearity is particularly true for the plot of σ_S, where the data exhibit a very strong approximately monotonic relationship between σ_S and crackle rating, but are clearly grouped along a curve rather than following the straight line with which linear regression would attempt to fit the data. Concerns regarding the nonlinear form of the relationship appropriately apply to σ_N as well. The growth of crackle rating with σ_S and σ_N is similar in form to a plot of the logarithm; consequently, we next try transforming the independent variable using the form

$$\log_{10}(a + x) \tag{2}$$

to determine if this is an appropriate transformation to address the nonlinearity in our model. The value of a in the model is determined by optimization of r^2, which effectively means that a has been chosen to make the model as linear as possible. With this transformation the portion of the variance explained by σ_N and σ_S rises substantially from $r^2 = 0.72155$ and $r^2 = 0.79646$, respectively, to $r^2 = 0.90042$ and $r^2 = 0.95724$. Thus both variables are potentially very strong descriptors of crackle once the nonlinearity in the relationship is taken into account. Additionally, when this modification is applied to $\mathrm{Sk}\{\partial p/\partial t\}$, the variance explained increases from 93.3% reported in the study of Gee *et al.*[5] to 95.6%, making σ_S and $\mathrm{Sk}\{\partial p/\partial t\}$ the two best single-variable predictors of crackling sound quality identified within this study when employing this transformation.

The log transformation procedure of course reintroduces one of the issues which this paper set out to address: that of fitting data using functional relationships that have a domain which fails to include and give reasonable values for all of possible values of the independent variable. While these data *happen* to be compatible with this nonlinear transformation, nothing prevents one of the independent variable values from being small enough that $a + x$ assumes a negative value causing $\log a + x$ to yield a nonsensical result. We could solve this by imposing a piecewise argument outside the logarithm so that the independent variable is expressed as

$$\left\{ \begin{array}{lr} 0, & \text{if } x \leq 10^{c/b-a} \\ b\log_{10}(a+x)+c, & \text{if } 10^{c/b-a} < x < 10^{\frac{50-c}{b}} - a \\ 50, & \text{if } 10^{\frac{50-c}{b}} - a < x, \end{array} \right\} \tag{3}$$

or we could impose a similar piecewise argument inside the logarithm so that the possibility of an out-of-bounds argument to the logarithm is foreclosed. Using the logistic function within the transformation could also keep all inputs in bounds and would have the additional benefit of maintaining derivative continuity.

C. LOGISTIC FUNCTION FIT TO DERIVATIVE SKEWNESS

Recalling the earlier success with reducing RMS error by fitting a logistic function to the derivative skewness, we again consider that form here. Because the logistic function better captures the particular form of the nonlinearity of the relationship of the logarithm of the derivative skewness with crackle rating, this model predicts 96.1% of the variance in the data, making it the best single-variable predictor identified in this study, as shown in the lower left panel of Fig. 4.

D. TWO-VARIABLE PREDICTION USING DERIVATIVE SKEWNESS AND STANDARD DEVIATION OF SHARPNESS

The two most descriptive metrics—the logistical fit to the logarithm of $\mathrm{Sk}\{\partial p/\partial t\}$, and the logarithmically transformed σ_S—were placed in a linear regression together, and the resultant fit is shown in Fig. 5, where 97.8% of the variance is explained by their linear combination.

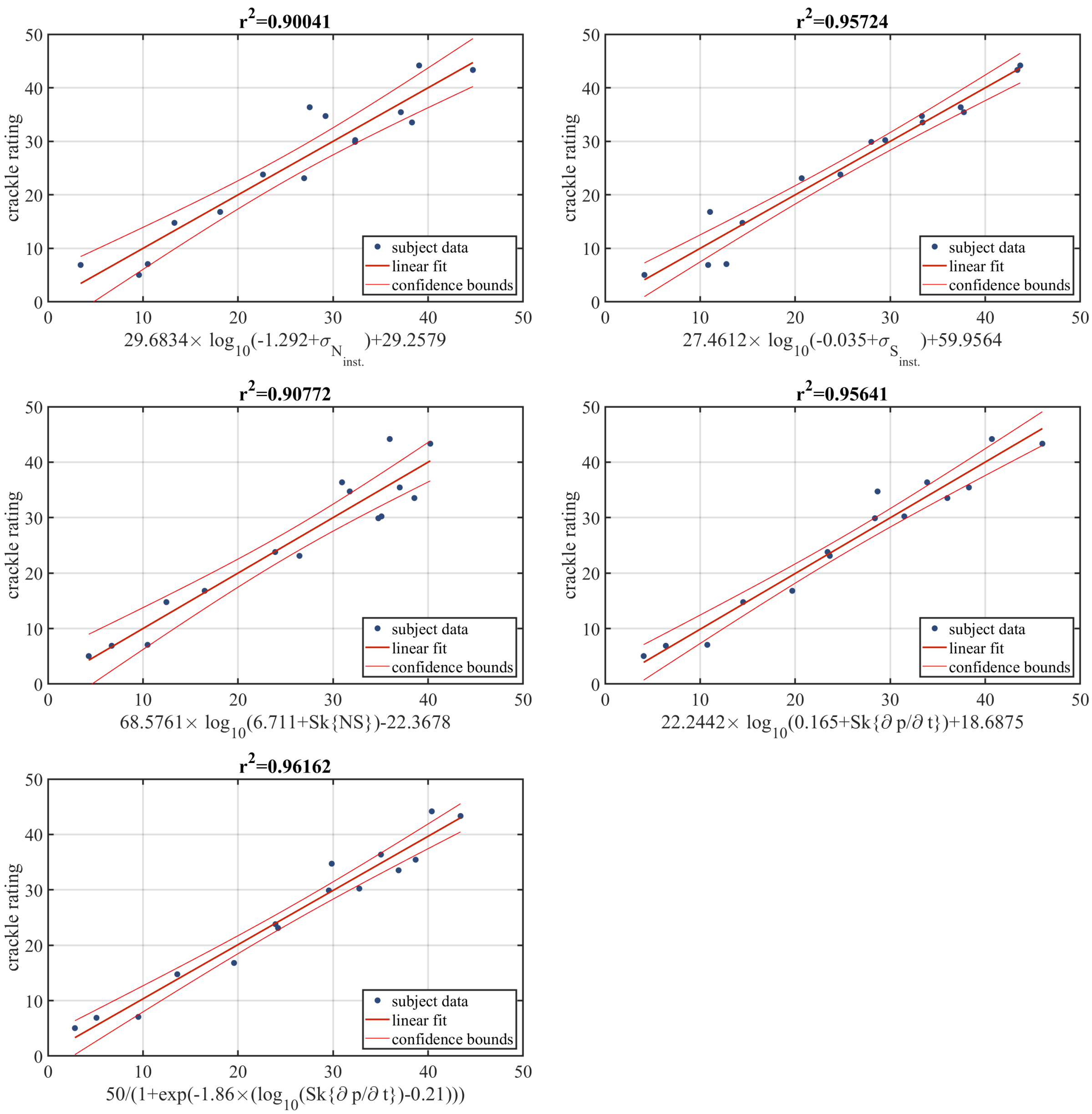

Figure 4: ***Raw subject crackle ratings shown as a function of each of the proposed measures with nonlinear adjustments as indicated in the x-axis labels. After applying a logarithm transformation to deal with their nonlinearity, the variance explained by*** σ_N ***and*** σ_S ***increases substantially, and the skewness-based measure's explanatory capacity increases marginally. When a logistic fit is applied to*** $\mathrm{Sk}\{\partial p/\partial t\}$***, 96% of variance is explained.***

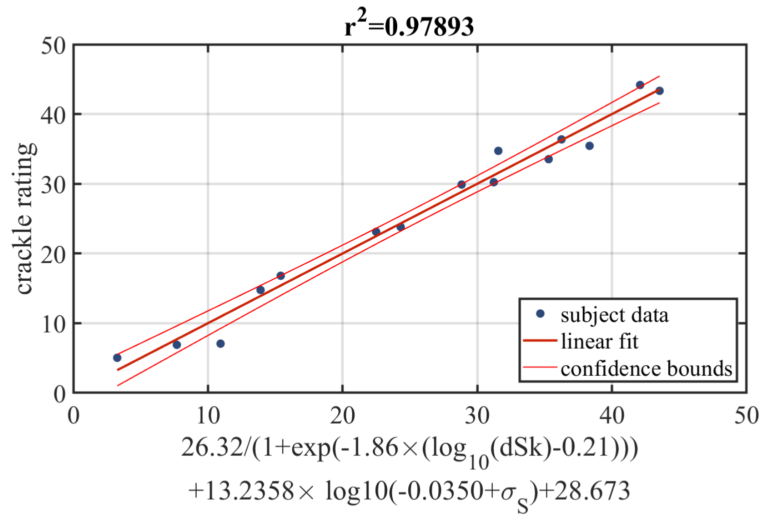

Figure 5: The logistic fit to logarithm of the derivative skewness and the $\log(a + \sigma_S)$ ***are used together in a linear regression to explain 97.89% of the variance.***

E. TRANSITIONAL FIT TO DERIVATIVE SKEWNESS AND SHARPNESS MEASURES

Because the logistic fit to the logarithm of the derivative skewness and the sharpness each have differing regions in which they are relatively poor predictors of crackling sound quality (the lower crackle values for sharpness, and the higher for derivative skewness) one could envision a blended function which performed better than either alone. This could, for example, be accomplished by expressing the two relationships under consideration as

$$a_1 = 27.4612 \times \log_{10}(-0.0350 + \sigma_S) + 59.9564,$$

and

$$a_2 = 50/(1 + \exp(-1.86 \times \log_{10}(\mathrm{Sk}\{\partial p/\partial t\} - 0.21)),$$

and using a logistic function

$$\phi = 1/(1 + exp(-0.288(a_2 - 21.94)))$$

to weight the two functions as

$$a_1 \times \phi + a_2 \times (1 - \phi).$$

When this approach is followed, the resultant prediction capitalizes on the seeming heteroscedasticity in each predictor by combining the low-scatter regions of each to obtain a relationship which predicts 98.87% of the variance as seen in Fig. 6.

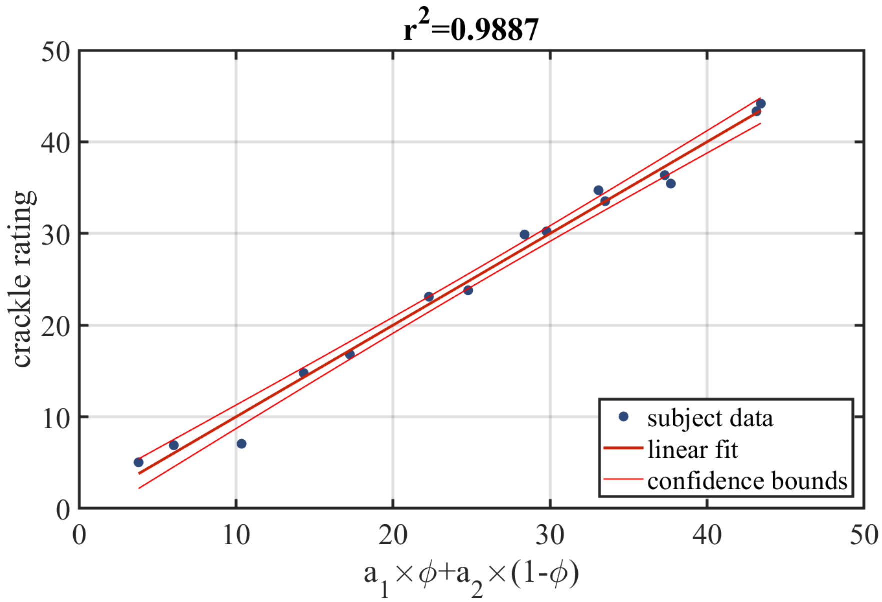

Figure 6: The logistic fit to logarithm of the derivative skewness and the $\log(a{+}\sigma_S)$ ***are combined together using a logistic blending function to explain 98.87% of the variance.***

However, at this point it is worthwhile to recall the statement famously attributed to Johnny von Neumann[11] that "with four parameters I can fit an elephant, and with five I can make him wiggle his trunk." Given the degree of optimization and the number of parameters involved in obtaining the final relationship, substantial caution should accompany any effort to apply this relationship outside of the present data set before rigorously testing it using a wider database of crackle rating data. Particularly, the qualitative result that $\mathrm{Sk}\{\partial p/\partial t\}$ performs better at low crackle ratings and σ_S performs better at high crackle ratings needs to be evaluated using data from other studies. For example, a larger study of crackle rating was performed at Brigham Young University in collaboration with Blue Ridge Research and Consulting. The data from that study would be a natural choice for evaluating the generalizability of the metrics identified in this study, and the predictive scope of the relationships identified between them and rated crackle.[12,13] Additionally, repeating this analysis with a greater degree of statistical rigor (for example using leave-one-out cross validation) could allow us to more easily assess whether these models are justified or suffer from overfitting. Furthermore, several of the runners up—σ_N, $\mathrm{Sk}\{\mathrm{NS}\}$, σ_N, ρ_{SN}—also had regimes for which they performed well; consequently, a variety of regime-specific combinations could be explored. Finally, the metrics that have performed well in previous studies but were not included in this study—in particular, roughness and sharp roughness—should be included in a definitive analysis.

4. CONCLUSION

We reanalyzed published subject response data from a crackle rating study by Gee *et al.*[5] by applying sound quality metrics to the original waveforms as well as a collection of transformations, including a logistic curve fit, to address nonlinearity between the dependent and independent variables. Gee *et al.*'s original linear fit accounted for 93.3% of the variance using $\mathrm{Sk}\{\partial p/\partial t\}$; this rose to 95.6% when using a transformation of the form $\log_{10}(a + bx)$. The standard deviation of the sharpness σ_S explained 79.6% of the variance as a single linear variable; however, this rose to 95.7% when a transformation of the form $\log_{10}(a + bx)$ was employed to deal with the functional relationship's nonlinearity. When a logistical function of $\mathrm{Sk}\{\partial p/\partial t\}$ was fit to the subject data, the portion of the variance explained rose to 96.1%. When the logistical function of $\mathrm{Sk}\{\partial p/\partial t\}$ was used in a linear regression with the standard deviation of the sharpness transformed using $\log_{10}(a + bx)$, the resultant linear combination explained 97.9% of the variance. A final fit taking advantage of the superior prediction quality from $\mathrm{Sk}\{\partial p/\partial t\}$ at low crackle levels and σ_S at higher was constructed using a logistic transition between the two, and this accounted for 98.9% of the variance; however, this last result effectively rests on generalizing the qualitative performance of the two component metrics within this study, and should be confirmed using data other than that originally used to devise it before applying it more generally. Additionally, future efforts should include leave-one-out cross validation to assess whether and to what degree overfitting may have occurred.

5. LINGERING QUESTIONS

Despite predicting all but a trivial portion of the variance in crackle ratings from the study of Gee *et al.*,[5] a few limitations exist. Although the study's signals were available for reanalysis, these were the signals as used to drive the loudspeaker rather than *as received by listeners*. Although the loudspeaker's amplitude frequency response was verified and acceptable, minor phase differences may exist between the output signal and the received signal and, if present, could have affected both the received derivative skewness as well as the sound quality metric values.

Additionally, given the minimal unexplained variance, small differences in loudness normalization—loudness significantly affects crackle perception—may explain some of the residual variance; however, without knowing the exact calibrations of the signals, this possibility is difficult to assess.

However, given the strength of the result, it is difficult to fret too much about whether this may have robbed us of some portion of the remaining 2.2% or 1.1% worth of conceivable variance, depending on which of the two final formulae one employs.

ACKNOWLEDGMENTS

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan.

REFERENCES

[1] J. E. Ffowcs Williams, J. Simson, and V. J. Virchis, "'Crackle': An annoying component of jet noise", *J. Fluid Mech.* **71**, 251-271 (1975). https://doi.org/10.1017/S0022112075002558

[2] A. Krothapalli, L. Venkatakrishnan, and L. Lourenco, "Crackle: A dominant component of supersonic jet mixing noise", AIAA Paper No. 2002-2024 (2000). https://arc.aiaa.org/doi/pdf/10.2514/6.2000-2024

[3] B. O. Reichman, M. B. Muhlestein, K. L. Gee, T. B. Neilsen, and D. C. Thomas, "Evolution of the derivative skewness for nonlinearly propagating waves", *J. Acoust. Soc. Am.*, **139**, 1390-1403 (2016). https://doi.org/10.1121/1.4944036

[4] S. H. Swift, K. L. Gee, and T. B. Neilsen, "Transformations of a crackling jet noise waveform and potential implications for quantifying the "crackle" percept", *Proc. Mtgs. Acoust.*, **22**, 045005 (2014). https://doi.org/10.1121/2.0000468

[5] K. L. Gee, P. B. Russavage, T. B. Neilsen, S. H. Swift, and A. B. Vaughn, "Subjective rating of the jet noise crackle percept", *J. Acoust. Soc. Am.* **144**, EL40-EL45 (2018). https://doi.org/10.1121/1.5046094

[6] P. B. Russavage, T. B. Neilsen, K. L. Gee, S. H. Swift, "Rating the perception of jet noise crackle", *Proc. Mtgs. Acoust.* **33**, 040001 (2018). https://doi.org/10.1121/2.0000821

[7] S. H. Swift, K. L. Gee, T. B. Neilsen, J. M. Downing, M. M. James, "Exploring the use of time-sensitive sound quality metrics and related quantities for detecting crackle", *Proc. Mtgs. Acoust.* **30**, 040003 (2017). https://doi.org/10.1121/2.0000544

[8] S. H. Swift, K. L. Gee, "Extending sharpness calculation for an alternative loudness metric input", *J. Acoust. Soc. Am.* **142**, EL549–EL554 (2017). https://doi.org/10.1121/1.5016193

[9] B. R. Glasberg, B. C. J. Moore. "A model of loudness applicable to time-varying sounds", *J. Aud. Eng. Soc.* **50**, 331-342 (2002). https://aes2.org/publications/elibrary-page/?id=11081

[10] S. H. Swift and K. L. Gee, "Examining the use of a time-varying loudness algorithm for quantifying characteristics of nonlinearly propagated noise (L)", *J. Acoust. Soc. Am.* **129**, 2753-2756 (2011). https://doi.org/10.1121/1.3569710

[11] F. Dyson, "A meeting with Enrico Fermi", *Nature* **427**, 297 (2004). https://doi.org/10.1038/427297a

[12] M. Calton, M. M. James, M. Downing, A. Vaughn, K. L. Gee, G. H. Wakefield, "Connecting high-power jet noise characteristics with human annoyance: Listener trials", *J. Acoust. Soc. Am.* **146**, 3043 (2019) https://asa.scitation.org/doi/abs/10.1121/1.5137545

[13] B. M. Harker, J. M. Downing, M. M. James, K. L. Gee, "Application of a crackle-based adjustment to military aircraft noise levels", *J. Acoust. Soc. Am.* **152**, A222 (2022) https://asa.scitation.org/doi/abs/10.1121/10.0016079