SAND2022-15796C

1. INTRODUCTION

Noise is a problem in hospitals. Restorative sleep is important to recovery and the sonic environment in hospitals—one complicated by the use of fairly assertive alarms to ensure that sometimes-overworked medical personnel become alert to emerging needs—is often not conducive to uninterrupted restful sleep.[1] However, it is one thing to know that a problem exists, another to quantify it, and another still to solve the problem. While each of these steps is important in its turn, this paper focuses on the second step: the question of how best to quantify hospital noises in ways relevant to the preservation (or disruption) of restful sleep. The current effort began when I happened upon an ASA presentation given by Jo Solet and learned that she and her partners for the study had measured subjects' sleep state transitions using polysomnography while they were exposed to various characteristic hospital noises recorded within a hospital environment, and administered in the laboratory to healthy subjects prescreened for intact hearing.[2,3] The characteristics of the original study were excellent in several regards including the use of polysomnography (the combination of electroencephalogram, electrooculogram, electrocardiogram, and electromyogram widely hailed as the gold standard in sleep state assessment[4]), and the use of a wide variety of hospital-relevant sounds with different distinguishing traits that could allow for a variety of informative comparisons.

After hearing the talk, I asked if they had applied sound quality metrics to their sounds or if they had used only the reported A-weighted measure. Sound quality metrics had not yet been applied and, more importantly, the sound files used in the original study were retained so it was possible to reanalyze the data in order to learn what additional insights these tools might make available.

2. DATA PREPARATION

The sound files were calibrated by assuming a maximum presentation level of 70 dB A-weighted overall level for all sounds. Note: The variable used in the original analysis was "A-weighted fast-response sound level" exceeded 10 percent of the time within each 10-second signal; however, this is the value used in the analysis rather than the values that were used to determine the presentation levels. From there, the sounds were attenuated in 5-dB increments to reproduce the variety of stimulus levels from the study, and the sounds were evaluated as discussed in the section 3.

The data from the middle row of Fig. 3 of Ref. 2 was captured by applying the MATLAB user-created function grabit[5] to screen captures of its three individual subplots, which allowed the underlying values to be retrieved with acceptable accuracy. The form of these data appears to be sigmoidal or, synonymously, logistic, which makes sense because this functional form captures the asymptotic behavior of the probability: at sufficiently low levels the sound would be imperceptible and be expected to elicit no additional awakening whatsoever (resulting in a lower arousal probability value approaching zero or no additional effect), and at extreme high levels the sounds would be expected to elicit effectively universal awakening in individuals with hearing intact (resulting in a probability value approaching one). Consequently, the data were transformed using the formula

$$p_T = -\log\left(\frac{1-p}{p}\right), \tag{1}$$

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

which as the inverse of the logistic function with known upper bound, 1, and lower bound, 0, transforms probabilities, p into roughly linear functions, y_T, of underlying variables; these transformed probabilities were then used in subsequent analyses. To gain an idea of the effect of this transformation, the real and transformed probability values are plotted in Fig. 1 for stage 2 and stage 3 sleep, N2 and N3, respectively. Note that, in contrast to the original probabilities, the transformed probabilities have nearly constant slope once an initial threshold value is achieved. A possible explanation for this threshold effect is that the growth of loudness is much more rapid at low levels as a result of cochlear amplification effects; potentially this deviation from linearity may be alleviated when loudness metrics are employed that take this effect into account (in contrast to level metrics which do not). Nevertheless, the near-constancy of slope holds both within individual lines and among collections of lines, making the transformed variable a nearly ideal input for linear regression.

3. SOUND QUALITY METRICS

The sounds from the original study (prepared as described in Sec. 2) were evaluated using an implementation[6] of the time-varying extension to the ANSI S3.4-2007 loudness standard,[7] and a compatible implementation of sharpness.[8] Because the loudness metric employed provides a sequential time series of loudness values spaced a 1-ms intervals,[9] some measure that aggregates or summarizes the values needs to be adopted. Several measures have been proposed as possibly capturing aspects of the loudness of a sound; these include: the short-term loudness for the loudness experienced instantaneously, and the peak of the long-term loudness for the overall loudness impression for an extended sound.[9] However, it is not immediately clear which measures would be the best available predictor of arousal from sleep. Accordingly, a basket of loudness-based measures were tried including maximum short-term loudness, the log of the short-term loudness rate of change, and the log sum of the loudness over the course of the signal (somewhat like a sound exposure level measure).

The sounds were also evaluated informally for tonality—the degree to which a sound is qualitatively similar to the subjective experience of a single-frequency sine wave—by the first author, dividing them into tonal, somewhat tonal and not tonal categories numerically represented as 1, 0.5, and 0, respectively, an admittedly crude approach. During the presentation of this study at 2022 Denver ASA Healthcare Acoustics session, hosted by the senior author, the sounds were played for the audience who rated them informally; the senior author participated in the raw audience poll, and I will include the results of this informal poll with the understanding that this interactive and entertaining activity made no serious effort to rise to the level of rigor of a formal listening study. The results of these informal activities are given in Table 1. In this table the first author ratings, the raw audience votes that a particular sound was "tonal," and the number of audience votes divided by the highest number of votes received are given in the second, third, and fourth columns, respectively, while the sound description is given in the first.

4. ANALYSIS

With sound quality metric values for loudness, sharpness, and informal listening observation measures for tonality, linear regression was performed on the transformed probabilities, p_T. Linear

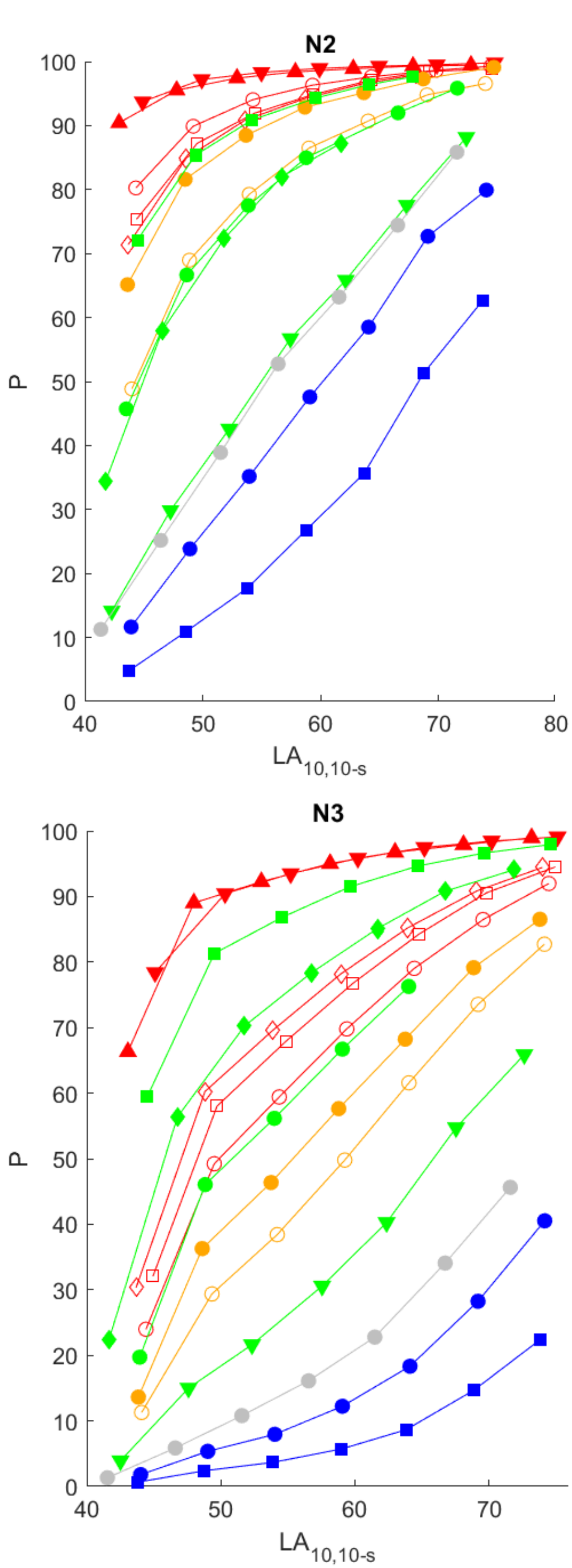

Figure 1: The use of Eq. 1 to transform the probability results in approximately linear behavior across its range after an initial minimum threshold is achieved. Note: Fig. 4 includes the legend for this figure.

Table 1: Informal tonality ratings by the first author, and by the talk audience

Sound	my rating	audience poll (raw)	audience poll (normalized)
Alarm	1	9	0.64
Phone ringing	1	14	1
Good conversation	0	1	0.07
Bad conversation	0	2	0.14
Paging	0	1	0.07
Door creak	0.5	6	0.42
Horrific copier	0.5	6	0.42
Snoring	0	1	0.07
Ice machine	0	5	0.36
Toilet flush	0	0	0
Laundry cart	0	0	0
Traffic?	0	1	0.07
Helicopter	0	3	0.21
Jet aircraft	0	0	0

regression attempts to determine the linear function that minimizes the residuals (deviations from the linear function within the data) to achieve the best possible fit. To begin with, the loudness or loudness-like measures are evaluated in order to select the one that explains the largest portion of the awakening probability. This will enable us to identify and focus on the most salient loudness-like parameter within the later analyses. The loudness-like measures under consideration are:

- maximum short-term loudness,
- log sum of short-term loudness,
- log maximum rate of change of short-term loudness, and
- $LA_{10,10-s}$ measure from original analysis.

The results were fairly unambiguous: as seen in Fig. 2; the sound exposure level-like log sum over the time history $\log \sum_T S$ explained only 6.6% of the variance, and could thus be excluded as the main loudness-based measure; the maximum short-term loudness, $\max S$, did somewhat better at 13%; however, neither of these did as well as the measure originally used in the paper, the fast A-weighted level exceeded ten percent of the time, $LA_{10,10-s}$, which predicted 34% of the variance; the log of the maximum rate of change of the short term loudness, however, did considerably better than all of the other measures accounting for 63.5% of the variance, a nearly two-fold boost in explained variance over the original measure. The log rate of increase of loudness is thus a far stronger predictor of arousal in at least this sleep stage than the other loudness-related factors examined.

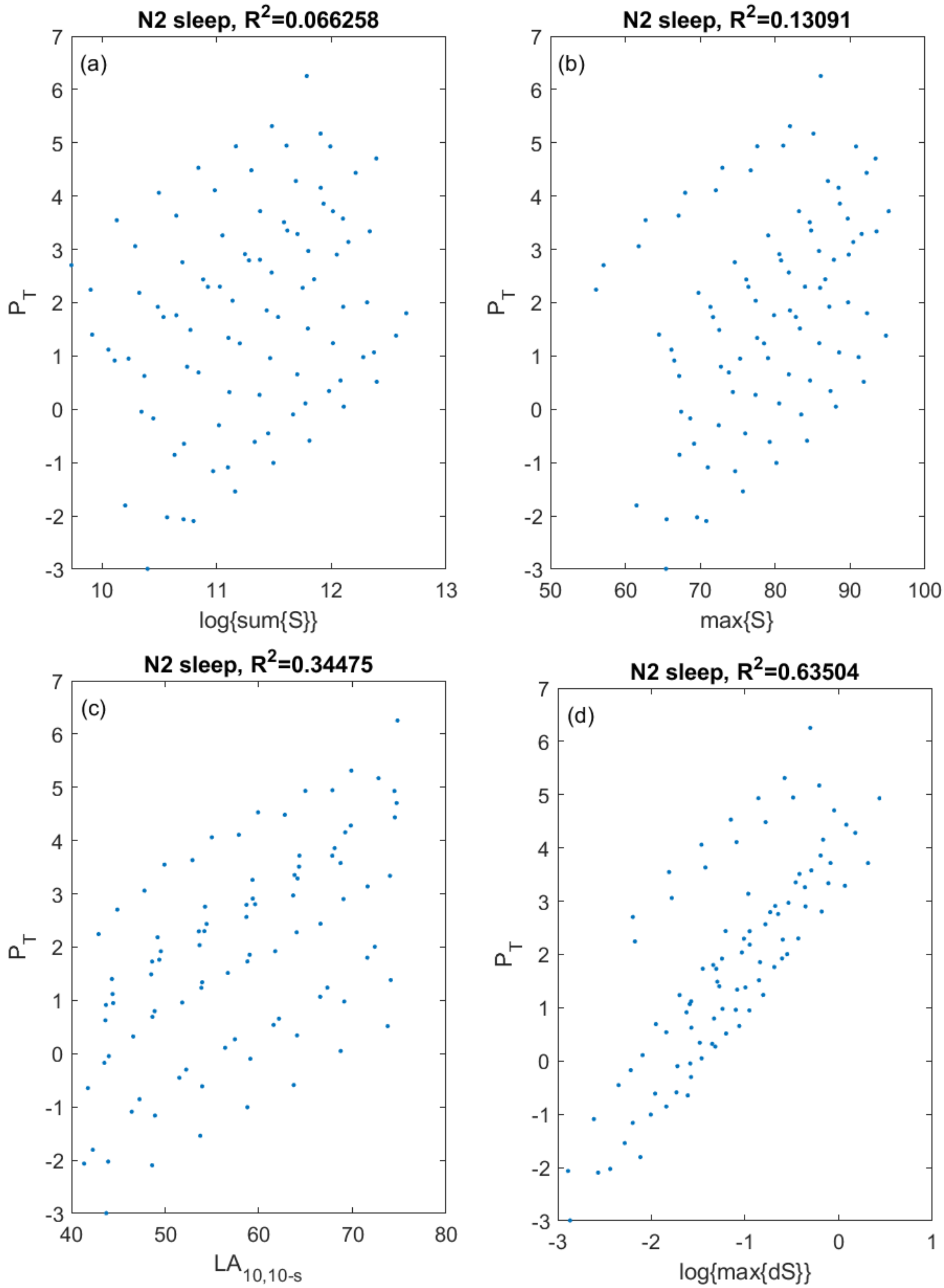

Figure 2: Of the four options explored, the rate of change of loudness (d) is far more effective at predicting arousal by hospital noises, explaining nearly twice as much of the variance of the transformed probability as the next most capable predictor, $LA_{10,10-s}$ ***(c).***

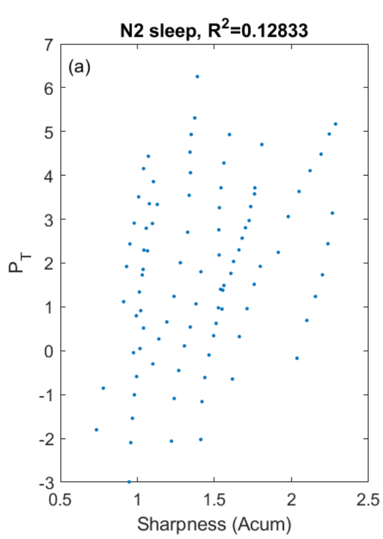

Figure 3: Both sharpness and tonality contribute to arousal; however, of the two, tonality appears to contribute more as a single variable, and the audience-assessed tonality was more effective at predicting arousal than the author-assessed version, which predicted nearly a third of the total variance!

The sharpness measure and the first author and audience-based ratings of tonality are next considered, as shown in Fig. 3. The sharpness measure explains 12.8% of the variance as a single-variable contributor. Of the two tonality ratings, the first-author rating of the tonality (part (b) of the figure), which categorized it into tonal, 1, somewhat tonal, 0.5, and not tonal, 0, predicted 28.8% of the variance; this was less than the audience-based measure (in which individual audience members voted for tonal, 1, or not tonal, 0, and the sum was normalized by the maximum number of votes received, which predicted 32.5% (nearly a third) of the variance as a single variable. It is important to realize that these additional metrics can be correlated to some finite degree with loudness and with each other; consequently, their total contributions to a combined model may differ somewhat from the sum of their individual shares of the variance. With that caveat, models using log of loudness increase with each of these variables are next considered. When $\log\max dS$ and maximum sharpness are used together in a model, 68.2% of the variance is explained an increase of 4.7%; this substantial reduction in explanatory power when sharpness is combined with the loudness measure suggests that at least part of sharpness's predictive power comes from its correlation with loudness, upon which it has a modest dependency. When loudness and tonality as rated by the audience are used in a model, 83% of the variance is explained, an increase of 19.5% from the use of $\log\max dS$ alone, a substantial increase; however, when tonality rated by the first author is instead used, 85.3% of the variance is explained, an increase of 21.8%. Taking the model with $\log\max dS$ and author-rated tonality, and adding sharpness to it explains 86.3% of the variance, an increase of only 1%, which suggests that sharpness may not be an especially important explanatory factor in arousal responses to the sounds in this data set.

To give a more visual idea of the usefulness of the model in expressing the relatively simple relationship between the metric parameters and arousal, the probability of arousal is shown as originally expressed as a function of $LA_{10,10-s}$ and as a linear combination of $\log\max dS$ and author-rated tonality is shown in Fig. 4 for Stage 2 and Stage 3 sleep. The same model materially improves collapse for data from both stages of sleep, suggesting that loudness rate of change and tonality are similarly important during both depths of non-REM sleep. This effectively gives a model that very reliably predicts the likelihood of patient awakening in response to hospital noise sound qualities during N2 and N3 stage sleep. Notably, the collapse is substantially farther to the right for N3, suggesting that this stage of sleep has a higher threshold of arousal, consistent with the notion of the deeper phases of sleep being more biologically protected. However, within the context of the existing model, this could largely be taken care of by using a constant to adjust for non-REM sleep stage.

Also notable is the fact that there is still a significant amount of lateral spread within both the N2 and N3 data; this suggests that application of further sound quality metrics such as, possibly, roughness or fluctuation, could result in further collapse. Both measures evaluate changes in the temporal envelope: the sensation of roughness is most sensitive to amplitude modulations between 15 and 300 Hz while fluctuation is sensitive to modulation frequencies below 30 Hz, and both measures have been shown to contribute to annoyance.[10] While predicted psychoacoustic annoyance and ability to generate awakenings may not be identical to one another, the hypothesis that qualities that increase likelihood of annoyance may increase likelihood of awakenings seems both a good starting point, and one that matches human experience. However, it also bears stating that within this data set some of the alarm sounds exhibit both tonality and a rough or fluctuating character, raising the possibility that a portion of the share of variance attributed to tonality in the present model could end up being attributed to one of these qualities in a more inclusive model.

5. CONCLUSIONS, IMPLICATIONS, AND FUTURE EFFORTS

Several key qualitative take-aways are available from the present examination; perhaps the most import of these is that it is not necessarily the loudness of a sound that generates awakenings, but the rate at which the loudness increases, accounting for more than three fifths of the variance. Next to this is the key role of tonality in the present model in predicting a further fifth of the variance. While future analyses with a larger suite of metrics may improve upon these results, and may subtly alter the balance of prediction between metrics, the ability of sound quality metrics to predict arousal during non-REM sleep with greater precision than standard level-based metrics warrants their use in wider contexts in sleep research and, in particular, calls for their use in evaluating the sonic acceptability of devices in the health care service environment.

If we were to take the results of this study as perfectly capturing the psychoacoustic determinants of awakening, the implications for policy would be that awakenings could be prevented through limiting the rate of rise and tone-like qualities of sounds. As it stands, however, these policy implications should be taken as hypotheses to be tested rather than established formula for preserving sleep. They are, nevertheless, hypotheses worth testing inasmuch as they potentially point the way to significantly better preservation of sleep than is possible using the guidance from current A-weighted metrics. It is also worth noting that some of the same qualities, e.g., tone-like character, that seem to predict awakening also are features of sounds designed to attract attention, and may be important to the ability of medical personnel to appropriately respond to them. Consequently, it may be easier in some cases to alter the loudness rise profile of alarm sounds while still maintaining their functionality as alarms, than to remove the tonal qualities. The Senior author,

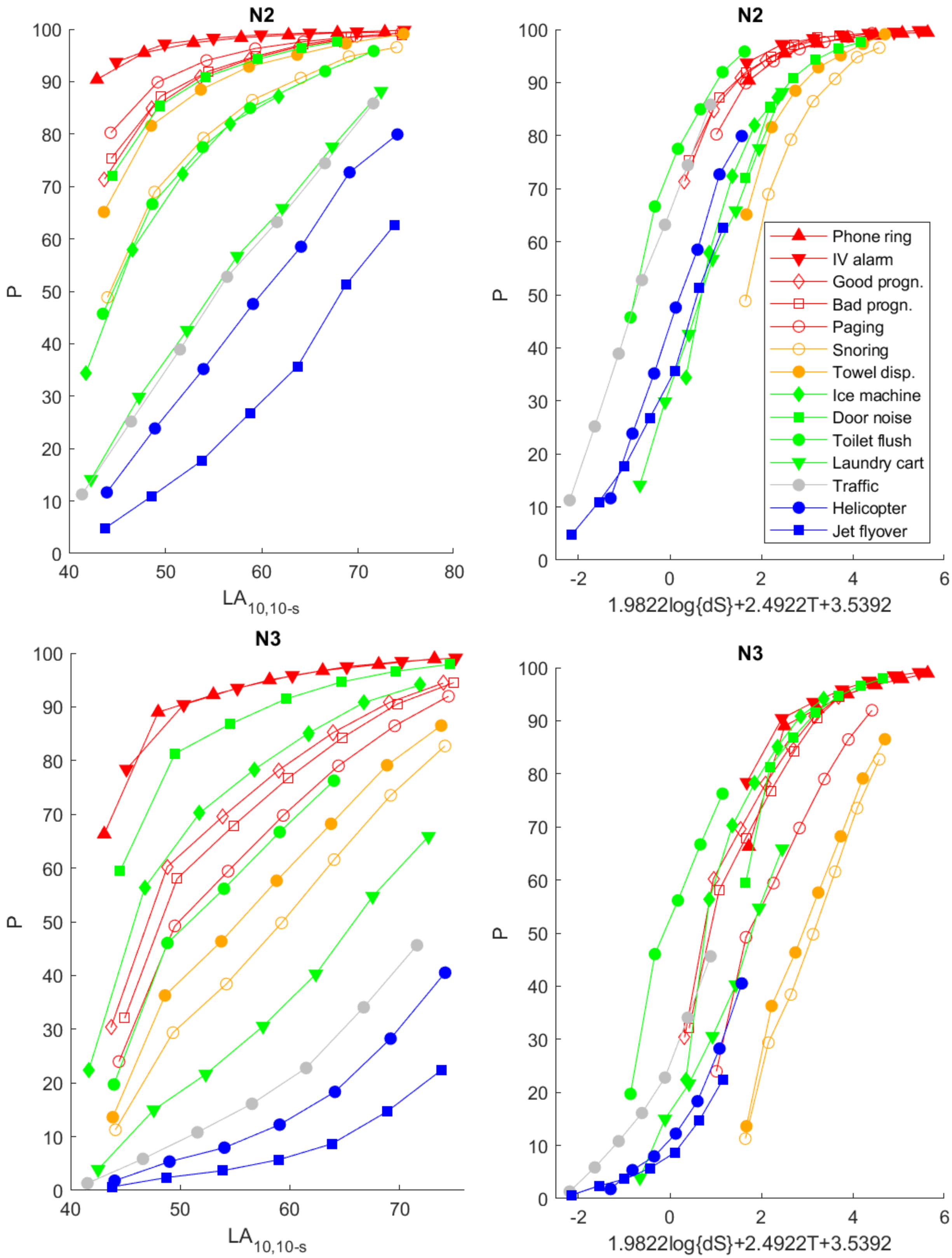

Figure 4:** The collapse of the data improves substantially when* $\log\max dS$ ***and author-rated tonality are considered. This is even true for stage 3 sleep for which the model was not originally adjusted.

who is trained as both a clinician and researcher, suggests an important follow-on study from a policy standpoint should be undertaken to develop a hierarchy of hospital alarm signals needed for sustained effectiveness in alerting medical personnel to immediate patient needs and to compare these for capacity to disrupt patient sleep. In a care environment crowded with a multitude of alarms with varying clinical significance or insignificance vie for attention,[1] options for limiting and streamlining alarms to focus attention on clinically relevant monitoring results should be explored and implemented. These efforts should be combined with existing exploratory efforts to preserve the sleep environment of hospital patients to maximize its restorative benefits.[11]

While this study has clearly opened the door to further study of the applicability of sound quality metrics to sleep disruption and demonstrated the power of this approach, much more remains to be done. This study has so far examined only two sound quality metrics and used informal evaluations of a third to evaluate the sounds. The informal evaluation should be replaced by either a jury study or a formalized tonality metric. Fluctuation and roughness should also be evaluated in order to remove the potential for confounding the effects of tonality with those of, e.g., roughness. Additionally, while this study has focused on NREM sleep stages completeness suggests that REM should be addressed also in order to determine whether the present model also has predictive benefits over level metrics for awakening during REM sleep.

ACKNOWLEDGMENTS

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Energy Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

REFERENCES

[1] Jo M. Solet, and Paul R. Barach, "Managing alarm fatigue in cardiac care," *Prog. Ped. Card.* **33**, 85-90 (2012).

[2] O. M. Buxton, J. M. Ellenbogen, W. Wang, A. Carballeira, S. O'Connor, D. Cooper, A. J. Gordhandas, S. M. McKinney, J. M. Solet, "Sleep disruption due to hospital noises: A prospective evaluation," *Ann. Int. Med.* **157**, 170-180 (2012).

[3] J. M. Solet, O. M. Buxton, J. M. Ellenbogen, W. Wang, A. Carballiera, "Evidence-based design meets evidence-based medicine: The sound sleep study." *The Center for Health Design, Concord, CA* (2010).

[4] M. Marino, Y. Li, M. N. Rueschman, J. W. Winkelman, J. M. Ellenbogen, J. M. Solet, H. Dulin, L. F. Berkman, O. M. Buxton, "Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared with polysomnography," *Sleep* **36**, 1-9 (2013).

[5] J. Doke (2003). GRABIT (https://www.mathworks.com/matlabcentral/fileexchange/7173-grabit), MATLAB Central File Exchange. Retrieved December 27, 2019.

[6] S. H. Swift (2020). Codes for the rapid calculation of loudness and sharpness (https://www.mathworks.com/matlabcentral/fileexchange/73808-codes-for-the-rapid-calculation-of-loudness-and-sharpness), MATLAB Central File Exchange. Retrieved July 29, 2022.

[7] S. H. Swift, and Kent L. Gee, "Techniques for the rapid calculation of the excitation pattern in the time varying extensions to ANSI S3. 4-2007." *Proc. of Meet. Acoust.* **36**, 040002 (2019).

[8] S. H. Swift, and K. L. Gee, "Extending sharpness calculation for an alternative loudness metric input," *J. Acoust. Soc. Am.* **142**, EL549-EL554 (2017).

[9] B. R. Glasberg, and B. C. J. Moore. "A model of loudness applicable to time-varying sounds," *J. Aud. Eng. Soc.* **50**, 331-342 (2002).

[10] H. Fastl, and E. Zwicker, "Psychoacoustics: Facts and Models, 3rd edition" *Springer* (2006).

[11] M. C. Bartick, X. Thai, T. Schmidt, A. Altaye, J. M. Solet, "Decrease in as-needed sedative use by limiting nighttime sleep disruptions by hospital staff," *J. Hosp. Med.* **5**, 20-24 (2010).