JASA ARTICLE

Analysis of correlation between window duration for kurtosis computation and accuracy of noise-induced hearing loss prediction

Yu Tian,[1,a)] Wenxi Ding,[1] Meibian Zhang,[2] Tianshu Zhou,[1] Jingsong Li,[1,b)] and Wei Qiu[3]

[1]*Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China*

[2]*Institute of Environmental and Occupational Health, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China*

[3]*Auditory Research Laboratory, State University of New York at Plattsburgh, Plattsburgh, New York 12901, USA*

ABSTRACT:
Kurtosis is considered an important metric for evaluating noise-induced hearing loss (NIHL). However, how to select window duration to calculate kurtosis remains unsolved. In this study, two algorithms were designed to investigate the correlation between window duration for kurtosis computation and the accuracy of NIHL prediction using a Chinese industrial database. Pure-tone hearing threshold levels (HTLs) and full-shift noise were recorded from each subject. In the statistical comparison, subjects were divided into high- and low-kurtosis groups based on kurtosis values computed over different window durations. Mann–Whitney *U* test was used to compare the difference in group HTLs to find the optimal window duration to best distinguish these two groups. In the support vector machine NIHL prediction model, kurtosis obtained from different window durations was used as a feature of the model for NIHL evaluation. The area under the curve was used to evaluate the performances of models. Fourteen window durations were tested for each algorithm. Results showed that 60 s was an optimal window duration that allows for both efficient computation and high accuracy for NIHL evaluation at test frequencies of 3, 4 and 6 kHz, and the geometric mean of kurtosis sequence was the best metric in NIHL evaluation.
https://doi.org/10.1121/10.0003954

(Received 6 July 2020; revised 22 January 2021; accepted 9 March 2021; published online 7 April 2021)

[Editor: Colleen G. Le Prell] Pages: 2367–2376

I. INTRODUCTION

Hearing loss is a major public health problem around the globe, and its negative impact is ranked high in the global disease burden (Nelson *et al.*, 2005). Occupational exposure to high-intensity noise is a leading cause of hearing loss: approximately 24% of the hearing problems among workers in the United States are caused by occupational noise exposure (Centers for Disease Control and Prevention, 2019). The ISO-1999 (2013) document is currently the most accepted noise damage/risk criterion. However, there has been much controversy regarding the ability of this criterion to estimate hearing loss caused by non-Gaussian complex noise (Zhao *et al.*, 2010; Xie *et al.*, 2016; Zhang *et al.*, 2020). According to the National Institute for Occupational Safety and Health (NIOSH) document (National Institute for Occupational Safety and Health, 1998), the ISO-1999 standard is based on experimental data of mixed quality and lacks data on the effects of temporal variables. Studies have shown that ISO-1999 is applicable for continuous or steady-state Gaussian noise (defined as noise that is normal-distributed) but that it underestimates noise-induced hearing loss (NIHL) caused by complex noise. A non-G noise consists of a background Gaussian noise that is punctuated by a temporally complex series of randomly occurring high-level noise transients. These transients can be brief high-level noise bursts or impacts (Ahroon *et al.*, 1993; Zhao *et al.*, 2010).

Studies show that kurtosis can effectively evaluate the biological effects of complex industrial noise and that it is a potential auxiliary index for evaluation (Davis *et al.*, 2009; Goley *et al.*, 2011; Davis *et al.*, 2012). Recent results from animal experiments have shown that for noise exposures having the same spectral energy, hearing trauma increases with increasing kurtosis, and kurtosis could be used to gauge the extent of NIHL (Lei *et al.*, 1994; Hamernik *et al.*, 2003; Qiu *et al.*, 2013). For a sample of n values, kurtosis is calculated as

$$\beta = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^2}, \quad (1)$$

where x_i is the ith value of the acoustic signal's amplitude, and $\overline{x}$ is the sample mean. Kurtosis is an index of the extent to which the distribution of a variable deviates from the

a)ORCID: 0000-0002-6791-8217.
b)Also at: Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, China. Electronic mail: ljs@zju.edu.cn.

Gaussian. Therefore, combining kurtosis with energy may predict NIHL more accurately.

The window duration has a large influence on the calculated kurtosis values, i.e., different kurtosis values are obtained when using different window durations. However, a method for selecting a proper window duration for kurtosis computation has not been studied thoroughly. It has been found that the kurtosis value is sensitive to the duration of the calculation window (Qiu *et al.*, 2006b; Smalt *et al.*, 2017); that is, different calculation window durations produce different kurtosis values, making the noise feature described by kurtosis somewhat ambiguous and potentially reducing the accuracy of evaluating NIHL by kurtosis. In the current studies, the most commonly used kurtosis calculation window duration was 40 s, and the kurtosis of the noise signal was computed over consecutive 40-s time windows (Hamernik *et al.*, 2003; Qiu *et al.*, 2006a; Davis *et al.*, 2009; Davis *et al.*, 2012; Xie *et al.*, 2016). This window duration was based on animal data (Hamernik *et al.*, 2003) and has not been verified with human workers exposed to occupational noises. Considering the diversity and complexity of industrial noise, it is necessary to test a larger range of window durations and to analyze the correlation between the kurtosis values obtained from these windows and the accuracy of the hearing loss assessment model to find the optimal window duration for computing kurtosis. In this study, the correlation between different time windows of kurtosis computation and the accuracy of the NIHL prediction model was explored based on a database containing data from subjects exposed to various types of complex noise.

II. MATERIALS AND METHODS

A. Data collection

1. Subjects

Industrial workers were recruited from 22 factories in the Zhejiang province of China between 2010 and 2017. The subjects ($N = 3242$) were introduced to the purpose and design of the study by an occupational physician and were asked to sign an informed consent form. The Zhejiang Provincial Center for Disease Control and Prevention (ZJCDC) institutional committee for the protection of human subjects approved the study protocol (approval reference number: ZJCDC-T-043-R).

2. Questionnaire survey

An occupational hygienist from ZJCDC administered a questionnaire to each subject to collect the following information: general personal information (age, sex, etc.), occupational history (e.g., factory, worksite, job description, length of employment, duration of daily noise exposure, and history of using hearing protection), personal life habits (e.g., smoking and alcohol use), and overall health conditions (including the history of ear disease and use of ototoxic drugs). An occupational physician entered all the information into a database.

3. Noise data collection

Shift-long noise recording files were obtained for each noise-exposed subject at the 22 factories using an ASV5910-R digital recorder (Hangzhou Aihua Instruments Co., Hangzhou, China). The ASV5910-R digital recorder is a specialized sound recording device that can be used for precision measurements and analysis of personal noise exposure. The instrument uses a 1/4-in. prepolarized condenser microphone characterized by good stability, high upper measurement limit, and wide frequency response (20 Hz–20 kHz). The microphone sensitivity is 2.24 mV/Pa, and the measurement range is 40–141 dBA. The microphone was mounted on the shoulder of the subject using special clips. The shift-long noise for each subject was recorded continuously by the ASV5910-R at a 32-bit resolution with a 48-kHz sampling rate. The noise recordings were saved on a 32-gigabyte micro secure digital (SD) card and transferred to network-attached storage for subsequent analysis.

4. Physical and audiometric evaluation

Each subject underwent a general physical and otologic examination. Air-conduction pure-tone hearing threshold levels (HTLs) at 0.5, 1.0, 2.0, 3.0, 4.0, 6.0, and 8.0 kHz were measured in each ear by an experienced audiologist. The tests were conducted in an audiometric booth [baseline noise $<$30 dB sound pressure level (SPL)] using an audiometer (Madsen, OB40; Otometrics, Copenhagen, Denmark) calibrated according to the Chinese national standard (GB4854-84). The audiograms were collected at least 16 h after the subjects' last occupational noise exposure.

B. Data inclusion

For inclusion in the study, all the subjects had to satisfy the following criteria: (1) consistently worked in the same job category and worksite (noise exposure area) for their entire career; (2) no history of genetic or drug-related hearing loss, head wounds, or ear diseases; (3) no history of military service, shooting, or setting off fireworks; and (4) no or minimal use of hearing protection (determined from the noise exposure questionnaire and interview). Accordingly, a total of 2110 workers were included from the original pool of 3242 subjects.

Two algorithms, a statistical comparison (algorithm 1) and a NIHL prediction model using the support vector machine (algorithm 2) were used to analyze the relationship between window duration and NIHL. The above-selected subjects were further screened to build cohorts for each algorithm according to the characteristics of these two algorithms. As a result, 1147 subjects were used in algorithm 1, while 847 subjects were used in algorithm 2. Figure 1 illustrates the data screening process of the two algorithms. The methods and justifications of data screening for these two algorithms are given below.

https://doi.org/10.1121/10.0003954

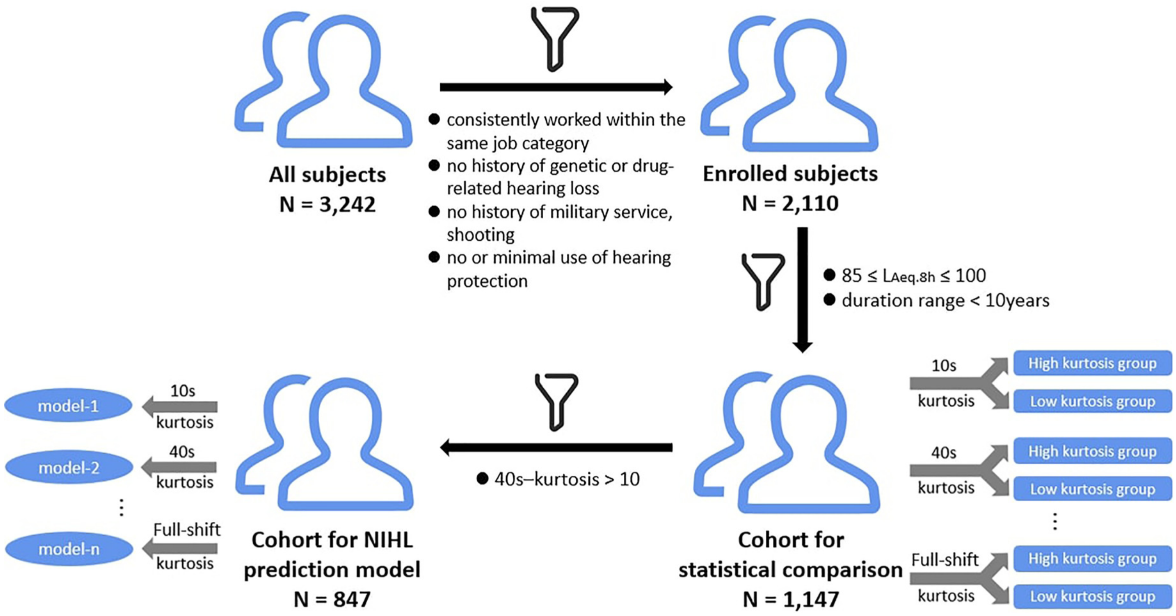

FIG. 1. (Color online) The overall sample screening and research process.

(a) Cohort for statistical comparison (algorithm 1): To focus on the relationship between kurtosis and hearing loss in the statistical comparison, the 2110 subjects were further screened according to the following conditions to build the statistical comparison cohort:

(1) 85 dBA $\leq L_{\mathrm{Aeq.8h}} \leq$100 dBA.
(2) Exposure duration of less than 10 yrs.

Condition (1) was set because NIHL began to occur obviously from 85 dBA (National Institute for Occupational Safety and Health, 1998), and the use of earplugs made it difficult to estimate the actual NIHL of workers exposed to noise greater than 100 dBA. Condition (2) was set because NIHL developed most rapidly in the first 10 yrs of exposure (Dobie, 2001; Zhang *et al.*, 2020). According to the above two inclusion conditions, a total of 1147 workers were included in the statistical comparison cohort.

(b) Cohort for NIHL prediction model (algorithm 2): There is some evidence that the kurtosis value is a sensitive measure when evaluating the NIHL of workers exposed to non-Gaussian noise (Xie *et al.*, 2016; Zhao *et al.*, 2019a). Thus, to highlight the correlation between the window used for kurtosis computation and NIHL, the workers exposed to non-Gaussian noise [defined as $\beta(\mathrm{t})_{40\mathrm{s}} > 10$] were screened out from the above 1147 workers in the statistical comparison cohort (details can be found in the supplementary material).[1] Consequently, 847 workers were enrolled in the cohort for the NIHL prediction model.

C. Definition of hearing loss

This study adopted two criteria for hearing loss:

(1) The NIOSH definition: the average HTLs at 1, 2, 3, and 4 kHz for both ears (denoted by HTL_{1234}) exceeding 25 dB HL. This is a widely used definition of work-related hearing loss.
(2) The average HTLs at 3, 4, and 6 kHz for both ears (denoted by HTL_{346}) exceeding 25 dB HL. This definition refers to frequencies that are most sensitive to NIHL evaluation (Dobie, 1995).

D. Kurtosis computation

The kurtosis of the recorded noise waveforms was computed over consecutive non-overlapping time windows over the full-shift noise records using a sampling rate of 48 kHz. To explore the relationship between the time windows and hearing loss, 14 different window durations were preselected (10 s, 40 s, 60 s, 2 min, 3.5 min, 4 min, 4.5 min, 5 min, 7.5 min, 10 min, 20 min, 30 min, 1 h, and full-shift). The number of kurtosis samples (time windows) ranged from 1 (full-shift 8 h) to 2880 (10-s window). How to obtain a single kurtosis value from this kurtosis value sequence that best describes the characteristics of the full-shift noise is one problem to be considered. In previous studies, both mean kurtosis (Xie *et al.*, 2016; Zhao *et al.*, 2019b) and median kurtosis (Zhao *et al.*, 2019a) have been used. In this study, three kurtosis metrics (i.e., mean, median, and geometric mean) were used to conduct a comprehensive analysis of the relation between the window duration and hearing loss prediction. Geometric mean was included as a kurtosis metric because, based on more than 2000 full-shift industrial noise recordings collected in China, kurtosis is distributed primarily in a lognormal manner (based on unpublished data).

E. Algorithm 1: Statistical comparison

As mentioned above, 14 window durations were tested for kurtosis computation. The kurtosis values were computed over each time window; then the mean, median, and geometric mean kurtosis values of kurtosis sequences were obtained. These three kurtosis metrics were used as criteria to divide the subjects into a low-kurtosis group and a high-kurtosis group. Under each kurtosis criterion, the quarter of the subjects with the lowest kurtosis value were partitioned into the low-kurtosis group, and the quarter of the subjects with the highest kurtosis value were partitioned into the high-kurtosis group (details can be found in the supplementary material).[1]

First, the differences in the hearing thresholds of the low- and high-kurtosis groups were analyzed for each time window. Then the results of each time window were compared to find the time windows over which the computed kurtosis could better distinguish the HTLs of the two groups. The Mann–Whitney U test was applied to analyze the difference between group HTLs. Multiple comparisons were corrected by the Šidák correction (Šidák, 1967).

However, in some large-scale assessments, even a small effect may reach statistical significance; consequently, the effect size was used to describe whether the effects have a relevant magnitude. Cohen's d (Cohen, 1988), which is not affected by sample size and is commonly used to measure effect size, was computed to measure the size of the differences in the average HTLs. In general, $d = 0.2$ indicates a small effect (the difference between the means of the two groups is small), $d = 0.5$ indicates a medium effect, and $d = 0.8$ indicates a large effect. Figure 2 shows the process of the statistical comparison algorithm.

F. Algorithm 2: Support vector machine (SVM) NIHL prediction model

According to recent studies, machine learning algorithms can be useful tools for predicting NIHL, and SVM is one of the most promising algorithms (Zhao *et al.*, 2019a; Zhao *et al.*, 2019b). An SVM is a supervised learning algorithm that constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space (Vapnik, 2000). SVM models can be used for classification, regression, or other tasks. In this study, it was used for classification (i.e., NIHL prediction, separating subjects into "NIHL" or "no NIHL").

The kurtosis metric computed over different time windows served as one of the features used to train SVM models to predict NIHL. Because age, gender, exposure duration, and equivalent A-weighted SPL (L_{Aeq}) are also significantly related to hearing loss (ISO-1999, 2013; Zhao *et al.*, 2019a; Zhao *et al.*, 2019b), these values also served as features (covariates) for the SVM models. All the features were normalized by Z-score, which is defined as

$$Z = \frac{x - \mu}{\sigma}, \qquad (2)$$

where x is the raw score, μ is the mean of the population, and σ is the standard deviation of the population. The area under the curve (AUC) was used to evaluate and compare the model performances to find suitable window durations that achieve better performance.

III. RESULTS

A. Descriptive statistical information

1. Basic information

Table I provides a breakdown of the average noise exposure level ($L_{\mathrm{Aeq.8h}}$), duration of exposure, age, and sex of the 1147 subjects enrolled in the statistical comparison cohort, corresponding to the number of subjects exposed in each plant. Most of the subjects were between 20 and 44 yrs old. The median age of the subjects was 31 yrs, and the mean was 33 yrs. The exposure durations for the 1147 workers ranged from 1 to 9.5 yrs. The median duration was 3 yrs;

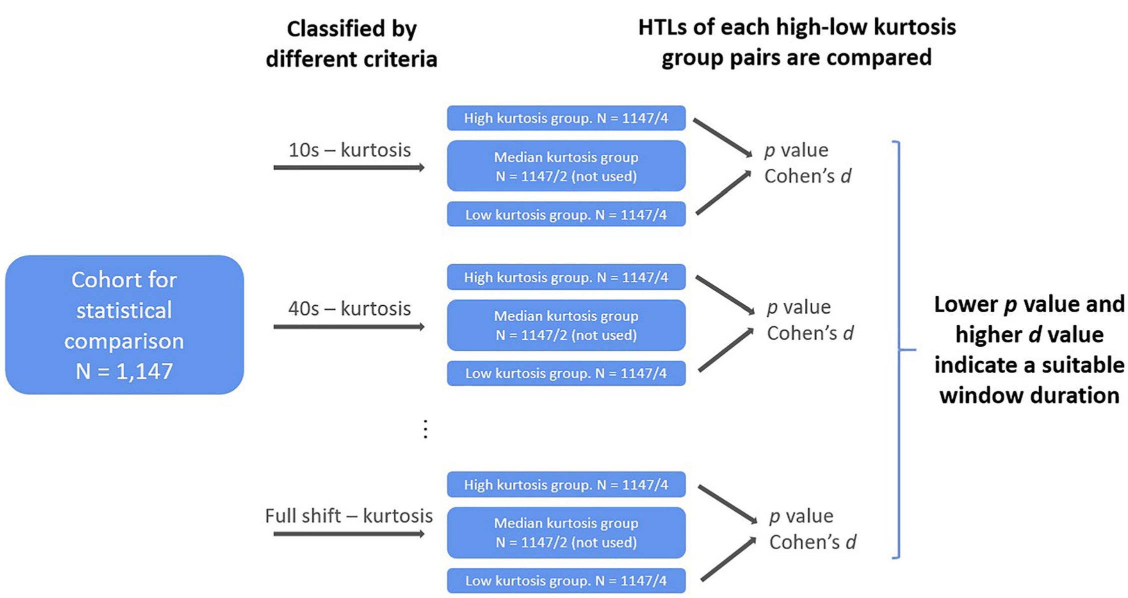

FIG. 2. (Color online) The process of the statistical comparison algorithm.

TABLE I. Descriptive statistical information on the characteristics of the 1147 subjects in the statistical comparison cohort in each factory (n, number of subjects in each group; ±, plus/minus 1 standard deviation; –, minimum to maximum).

Factory	Male (n)	Female (n)	Age (yrs)	Duration (yrs)	$L_{Aeq.8h}$ (dBA)
Textile mill 1	8	21	23–51 34.9 ± 5.58	2–9 6.86 ± 1.68	85.94–99.97 91.52 ± 3.81
Textile mill 2	8	73	17–46 28.94 ± 7.15	1–9 3.19 ± 2.29	93.31–99.63 96.98 ± 1.23
Textile mill 3	47	6	20–48 31.36 ± 7.6	2–9 4.55 ± 2.05	85.17–99.99 93.62 ± 4.16
Machinery plant 1	13	42	22–45 34.15 ± 6.47	1–9 4.87 ± 2.73	85.03–98.03 88.29 ± 2.8
Machinery plant 2	2	0	47–54 50.5 ± 3.5	7–8 7.5 ± 0.5	90.83–98.59 94.71 ± 3.88
Machinery plant 3	12	0	24–52 35.42 ± 9.01	2–9 6.17 ± 2.11	87.43–97.97 90.99 ± 2.88
Machinery plant 4	31	0	20–69 41 ± 12.22	1–9 5.42 ± 2.2	85.21–99.48 92.37 ± 3.94
Pipe manufacturing company	4	0	25–39 31 ± 5.1	3–9 6.75 ± 2.28	88.12–92.91 90.63 ± 1.81
Furniture factory 1	57	1	20–45 27.6 ± 5.26	1–7 2.43 ± 1.68	85.01–93.86 88.41 ± 2.09
Furniture factory 2	42	8	21–54 37.18 ± 8.51	1–9 4.96 ± 2.17	85.04–97.81 91.55 ± 3.06
Furniture factory 3	44	0	18–48 32.48 ± 8.18	1–9 4.23 ± 2.27	86.03–97.17 90.32 ± 2.92
Furniture factory 4	51	10	20–52 32.23 ± 7.65	1–9 2.23 ± 1.53	85.15–98.12 90.07 ± 3.12
Furniture factory 5	44	6	19–62 33.94 ± 9.88	1–9 3.06 ± 2.15	85.56–98.79 91.05 ± 2.44
Furniture factory 6	21	5	20–55 39.04 ± 11.3	1–9 4.35 ± 2.04	85.1–93.99 87.69 ± 2.19
Vehicle factory 1	244	30	19–55 32.42 ± 7.68	1–9 4.25 ± 2.4	85.13–98.05 89.11 ± 2.51
Vehicle factory 2	78	35	19–48 30.42 ± 6.96	1–9 2.8 ± 1.76	85.11–99.31 90.03 ± 3.27
Hardware factory	25	5	19–43 30.63 ± 6.77	1–9 3.9 ± 2.51	85.48–93.51 89.84 ± 2.45
Electronic equipment factory 1	33	7	17–43 26.75 ± 4.98	1–8 2.3 ± 1.78	86.3–99.78 91.09 ± 2.94
Electronic equipment factory 2	16	2	19–31 24.56 ± 3.29	1–7 2.67 ± 2	85.13–94.61 87.86 ± 2.14
Paper plant 1	14	7	20–61 42.62 ± 11.34	1–9 3.9 ± 2.51	85.06–96.94 90.2 ± 3.55
Paper plant 2	12	6	25–60 40.67 ± 10.12	1–9 5.67 ± 2.26	85.02–94.57 90.08 ± 2.96
Stroller factory	43	34	23–52 38.87 ± 8.1	1–9 3.05 ± 2.42	85.73–99.67 93.11 ± 3.32

the mean was 3.8 yrs. The median noise exposure level ($L_{Aeq.8h}$) was 90.1 dBA, and the mean was 90.7 dBA.

The breakdown of the average noise exposure level, duration of exposure, age, and sex of the 847 subjects enrolled in the NIHL prediction cohort can be found in Table S1 in the supplementary material.[1]

2. Kurtosis information

Based on a full-shift (8 h) noise record, kurtosis sequences of different lengths were obtained over different window durations. The analysis results show that most of the kurtosis sequences followed a lognormal distribution; however, a few obeyed a normal distribution. Based on these distribution characteristics, the following measures of the central tendency of the distribution were considered kurtosis metrics: median, mean, and geometric mean.

Figure 3 shows the changes in the average median, mean, and geometric mean kurtosis values in this study under different window durations. Clearly, as the window duration increases, the median, mean, and geometric mean kurtosis values increase significantly. In other words,

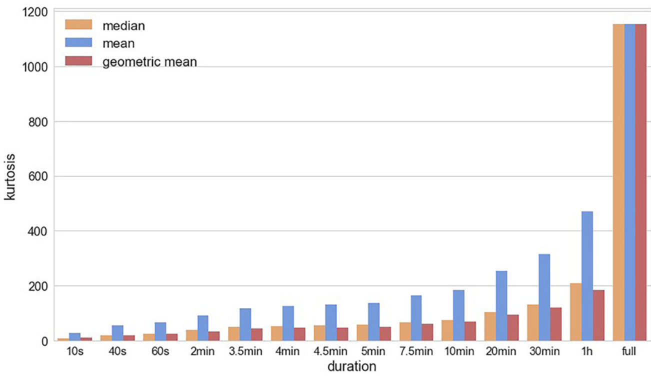

FIG. 3. (Color online) The changes in the average median, mean, and geometric mean of the kurtosis measurements under different window durations.

changing the window duration had a direct impact on the kurtosis value and consequently affected the relationship between the kurtosis metric and hearing loss evaluation. In addition, the median, mean, and geometric mean kurtosis values are significantly different ($p < 1 - (1 - 0.01)^{1/m}$, using the Šidák correction, $m = 3$, corresponding to the pairwise comparison of the three metrics) for all window durations except the median and geometric mean kurtosis computed over the 1-h window ($p = 0.039$).

B. Results for statistical comparison

Table II shows the difference in the average HTLs at 1, 2, 3, and 4 kHz (HTL_{1234}) between the high/low-kurtosis groups. The results of the three kurtosis metrics (i.e., median, mean, and geometric mean) are listed in the table. The p values were obtained by Mann–Whitney U tests. The top three Cohen's d values for each HTL difference are shown in boldface. Higher Cohen's d values indicate a larger difference in HTLs between the two groups. All the differences were significant for the compared groups ($p < 1 - (1 - 0.01)^{1/14} = 0.000\,72$) except for the pairs based on full-time kurtosis for the three metrics and the pairs based on 20 min and 1 h for the mean kurtosis. For median kurtosis, the p values of the 40-s to 3.5-min windows were less than 1×10^{-5}, and the effect sizes of the 40-s and 2-min windows were relatively large ($d > 0.4$). For mean kurtosis, the 2-min window had the lowest p value and the largest effect size. For geometric mean kurtosis, the p values of the 10-s to 2-min windows were all less than 1×10^{-5}, and the effect sizes of the 60-s to 3.5-min windows were the largest. Additionally, it could be seen that the top Cohen's d value corresponding to geometric mean kurtosis is the highest, and that corresponding to median kurtosis is the lowest.

Table III shows the difference in the average HTLs at 3, 4, and 6 kHz (HTL_{346}) between the high/low-kurtosis groups. All the differences were significant for the compared groups. For median kurtosis, the p values of the 40-s to 4.5-min windows were less than 1×10^{-9}, and the effect size of the 40-s window reached 0.646, indicating a relatively large effect. For mean kurtosis, the 2-min window had the lowest p value and the largest effect size. For geometric mean kurtosis, the p values of the 40-s to 7.5-min windows were less than 1×10^{-9}, and the effect size of the 60-s window reached 0.73, which was the largest effect. Similarly, the top Cohen's d value corresponding to geometric mean kurtosis is the highest, and that corresponding to median kurtosis is the lowest.

C. Results for the NIHL prediction model

The AUCs of the SVM models with kurtosis computed over different time windows are shown in Fig. 4. The performance of a model without kurtosis as an input feature (i.e., only age, $L_{Aeq,8h}$, and exposure duration were used as input features of the model) was used as a control to test whether kurtosis could improve the performance of NIHL prediction. AUC is the area under the receiver operator characteristic (ROC) curve, and its value ranges from 0 to 1. An area close to 1 indicates that the performance of the model is good,

TABLE II. The difference in the average HTLs at 1, 2, 3, and 4 kHz (HTL_{1234}) between the high/low median, mean, and geometric mean kurtosis groups.

	Median		Mean		Geometric mean	
Kurtosis window	p value	Cohen's d	p value	Cohen's d	p value	Cohen's d
10 s	7.51×10^{-5}	0.310	1.66×10^{-4}	0.371	3.73×10^{-6}	0.422
40 s	2.46×10^{-6}	**0.404**	2.80×10^{-4}	0.375	4.48×10^{-6}	0.434
60 s	2.35×10^{-6}	**0.394**	2.43×10^{-4}	0.379	3.59×10^{-6}	**0.460**
2 min	1.95×10^{-6}	**0.405**	3.92×10^{-5}	**0.442**	5.17×10^{-6}	**0.489**
3.5 min	8.48×10^{-6}	0.373	3.34×10^{-4}	0.380	3.05×10^{-5}	**0.472**
4 min	2.12×10^{-5}	0.359	3.12×10^{-4}	0.318	1.44×10^{-5}	0.457
4.5 min	2.06×10^{-5}	0.365	1.94×10^{-4}	0.345	1.07×10^{-5}	0.448
5 min	2.45×10^{-5}	0.366	8.88×10^{-4}	0.354	2.35×10^{-5}	0.444
7.5 min	2.48×10^{-5}	0.365	2.95×10^{-4}	**0.386**	1.35×10^{-5}	0.395
10 min	7.77×10^{-5}	0.345	1.44×10^{-4}	**0.382**	3.86×10^{-5}	0.352
20 min	2.05×10^{-5}	0.360	0.00174	0.343	3.78×10^{-5}	0.420
30 min	2.99×10^{-4}	0.294	1.32×10^{-4}	0.371	9.18×10^{-5}	0.369
1 h	4.48×10^{-5}	0.313	0.00438	0.225	2.04×10^{-4}	0.407
Full-time	0.00423	0.157	0.00423	0.157	0.00423	0.157

TABLE III. The difference in the average HTLs at 3, 4, and 6 kHz (HTL_{346}) between the high/low median, mean, and geometric mean kurtosis groups.

	Median		Mean		Geometric mean	
Kurtosis window	*p* value	Cohen's *d*	*p* value	Cohen's *d*	*p* value	Cohen's *d*
10 s	7.01×10^{-8}	0.482	3.51×10^{-8}	0.637	1.33×10^{-9}	0.670
40 s	2.11×10^{-10}	**0.646**	1.94×10^{-8}	0.639	3.07×10^{-10}	0.700
60 s	2.45×10^{-10}	0.595	2.97×10^{-8}	**0.650**	2.02×10^{-10}	**0.733**
2 min	6.18×10^{-11}	0.566	3.99×10^{-9}	**0.715**	3.10×10^{-10}	0.713
3.5 min	2.65×10^{-10}	0.553	2.80×10^{-7}	**0.656**	1.24×10^{-9}	**0.714**
4 min	7.99×10^{-10}	0.616	8.64×10^{-8}	0.594	3.67×10^{-10}	0.699
4.5 min	5.42×10^{-10}	0.607	8.72×10^{-8}	0.597	5.51×10^{-10}	0.696
5 min	1.99×10^{-9}	**0.639**	3.59×10^{-7}	0.621	7.94×10^{-10}	**0.715**
7.5 min	9.06×10^{-10}	0.605	1.33×10^{-7}	0.638	5.85×10^{-10}	0.625
10 min	3.99×10^{-9}	0.554	1.87×10^{-8}	0.621	1.84×10^{-9}	0.603
20 min	1.07×10^{-9}	0.503	1.05×10^{-6}	0.600	3.65×10^{-9}	0.667
30 min	2.80×10^{-8}	0.571	2.42×10^{-8}	0.608	3.73×10^{-8}	0.621
1 h	2.22×10^{-9}	**0.627**	4.93×10^{-6}	0.449	3.71×10^{-8}	0.670
Full-time	3.00×10^{-6}	0.346	3.00×10^{-6}	0.346	3.00×10^{-6}	0.346

while an area below 0.5 means that the performance is poor. In this study, the model performance was obtained by averaging 100 samplings that used 80% of data for the training set and 20% for the test set each time.

Figures 4(A) and 4(B) show the AUCs of the SVM models with *median* kurtosis. For the NIHL at 1–4 kHz, the model trained with kurtosis computed over a 40-s window had the highest AUC and was significantly higher than that of the model trained without kurtosis ($p < 0.0001$) based on the Mann–Whitney *U* test. For the NIHL at 3–6 kHz, the model trained with kurtosis computed over a 10-min window had the highest AUC, followed by the models with 40-s, 7.5-min, and 20-min windows. The AUC of these models was significantly higher than that of the control group. It can be seen that model performance degrades when the window duration exceeds 30 min.

Figures 4(C) and 4(D) show the AUCs of the SVM models with *mean* kurtosis. For NIHL at 1–4 kHz, the 10-s window model had the highest AUC, followed by the 40-s and 60-s window models. For the NIHL at 3–6 kHz, the 40-s window model performed best, followed by the models with 10-s and 60-s window durations. The AUCs of all the models mentioned above were significantly ($p < 0.0001$) higher than that of the model without kurtosis. It appears that model performance trends downward as the window duration increases.

Figures 4(E) and 4(F) show the AUCs of the SVM models with *geometric mean* kurtosis. For NIHL at 1–4 kHz, the model with a 40-s window had the highest AUC, followed by the models with 10- and 60-s window durations. For the NIHL at 3–6 kHz, the model with a 60-s window performed best, followed by the models with 40- and 10-s windows. The AUCs of all the models mentioned above were significantly ($p < 0.0001$) higher than that of the model without kurtosis.

IV. DISCUSSION

The earliest research on the relationship between kurtosis level and hearing loss dates back to 1994 (Lei *et al.*, 1994). Based on animal (chinchilla) data, Hamernik *et al.* (2003) found that a 40-s time window with a sampling rate of 48 kHz was sufficient for establishing an acceptable measure of kurtosis. However, the suitability of this selection for the evaluation of NIHL in humans has not been verified. In this study, the correlation between the window duration for kurtosis computation and the effectiveness of NIHL prediction was studied over two algorithms.

The statistical comparison (algorithm 1) was designed, since recent results from animal experiments (Hamernik *et al.*, 2003; Qiu *et al.*, 2006a; Qiu *et al.*, 2013) and epidemiological studies (Zhao *et al.*, 2010; Xie *et al.*, 2016; Zhang *et al.*, 2020) indicated that the HTLs of workers exposed to high-kurtosis noise should be significantly higher than those of workers exposed to Gaussian noise or low-kurtosis noise. Therefore, one can expect that the high- and low-kurtosis groups distinguished by an appropriate window duration will have more obvious HTL differences. The Mann–Whitney *U* test and Cohen's *d* were used to compare the HTLs of high/low groups based on different time window durations for kurtosis computation. The NIHL prediction models (algorithm 2) were built, since by comparing the performances of the models based on kurtosis computed over different windows, a better window duration could be obtained, because the kurtosis value computed over a suitable window would improve the predicting ability of the corresponding SVM model. The performances of the models were compared using their AUCs.

Furthermore, two hearing loss criteria (i.e., HTL_{1234} and HTL_{346}) and three kurtosis metrics (i.e., median, mean, and geometric mean) were used to ensure a comprehensive comparison.

A. Window duration selection via two algorithms

Johnson and Lowe (1979) indicated that for any sample of size *N*, the value of sample kurtosis increases as the sample size increases, and sample kurtosis is bounded by *N*. The

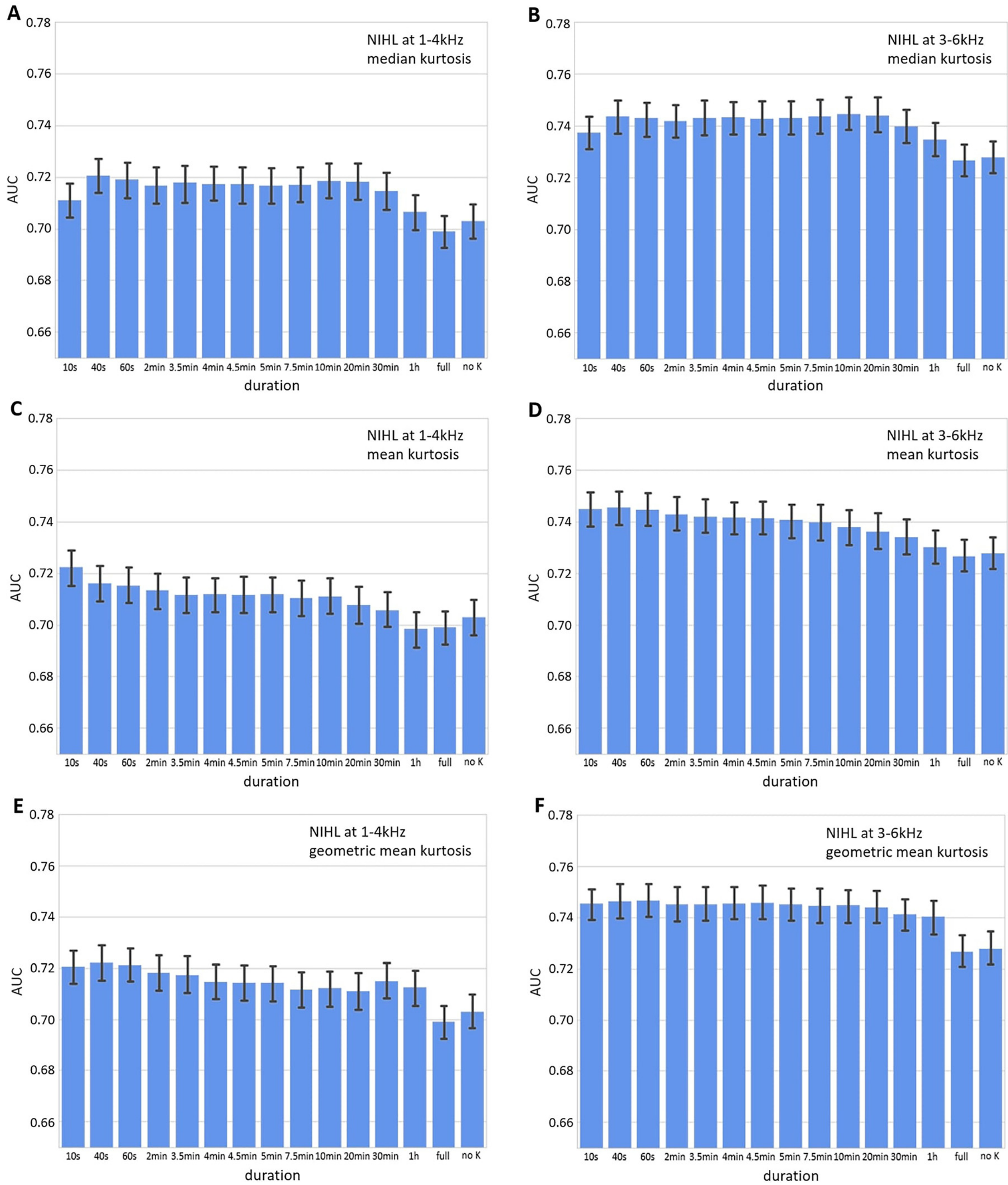

FIG. 4. (Color online) The AUCs of the SVM models for NIHL prediction with kurtosis computed over different window durations. (A) models for average HTLs at 1, 2, 3, and 4 kHz (HTL_{1234}) and median kurtosis; (B) models for average HTLs at 3, 4, and 6 kHz (HTL_{346}) and median kurtosis; (C) models for HTL_{1234} and mean kurtosis; (D) models for HTL_{346} and mean kurtosis; (E) models for HTL_{1234} and geometric mean kurtosis; (F) models for HTL_{346} and geometric mean kurtosis.

results of this study also show that, at a fixed sampling rate, as the sample size increases with the increase in the calculation window length, the kurtosis value increases (as shown in Fig. 3). The results of the two algorithms showed that kurtosis calculated with short window durations achieved better performance in NIHL evaluation than that calculated with long window durations. A short computation window duration makes the calculation more efficient and can timely track the dynamic characteristics of complex noise exposure.

As shown in Tables II and III, the difference in p and d between adjacent window lengths could be quite small (e.g., between 40- and 60-s windows), and the error bars between them overlapped, which means that it is difficult to determine which window is better only from a single comparison of an algorithm. However, the best window length can be selected by comprehensively evaluating the performance of each window length based on the results of two algorithms with two hearing impairment definitions and three kurtosis

TABLE IV. The top three window durations for each algorithm for average HTLs at 1, 2, 3, and 4 kHz (HTL_{1234}) and 3, 4, and 6 kHz (HTL_{346}). The top three selections for algorithm 1 were obtained by selecting the three largest Cohen's *d* values at HTL_{1234} and HTL_{346}; the top three selections for algorithm 2 were obtained by selecting the three largest areas under curves (AUCs) from the support vector machine (SVM) prediction models at HTL_{1234} and HTL_{346}.

	Median	Mean	Geometric mean
Algorithm 1 at HTL_{1234}	2 min, 40 s, 60 s	2 min, 7.5 min, 10 min	2 min, 3.5 min, 60 s
Algorithm 1 at HTL_{346}	40 s, 5 min, 1 h	2 min, 3.5 min, 60 s	60 s, 5 min, 3.5 min
Algorithm 2 at HTL_{1234}	40 s, 60 s, 10 min	10 s, 40 s, 60 s	40 s, 60 s, 10 s
Algorithm 2 at HTL_{346}	10 min, 20 min, 7.5 min	40 s, 10 s, 60 s	60 s, 40 s, 10 s

metrics (i.e., a total of 12 conditions). It is feasible to obtain a proper range for the window length or a suitable window length that corresponds to a smaller *p*, larger *d*, and better NIHL predicting performance taking all comparisons together. To facilitate the selection of the best window duration, the top three window durations obtained from algorithm 1 (using Cohen's *d*) and from algorithm 2 (using AUC) are summarized in Table IV. As shown in Table IV, the 2-min window appeared most frequently in algorithm 1, followed by window durations of 60 and 40 s. In algorithm 2, the 40-s window and 60-s window achieved the best performance in the NIHL evaluations for both the 1–4 and 3–6 kHz frequency ranges. Based on the above observation, the candidates for the best window duration for kurtosis computation are 40 s, 60 s, and 2 min (they appear at least four times in Table IV). Among them, the 60-s window appeared most frequently (9 of 12 conditions). This window duration performs well in both algorithms (appears four times in algorithm 1 and five times in algorithm 2), in both hearing loss definitions (appears five times at HTL1234 and four times at HTL346), and in all three metrics (appears two, three, and four times at median, mean, and geometric mean, respectively), indicating that it adapts to various conditions. Therefore, it is recommended to use 60 s as the window length to calculate kurtosis.

B. Kurtosis metric comparison

In this study, three kurtosis metrics (i.e., median, mean, geometric mean) were used, and their effects on HTL differences in algorithm 1 and NIHL prediction in algorithm 2 were investigated. In algorithm 1, the largest Cohen's *d* values obtained by geometric mean kurtosis, mean kurtosis, and median kurtosis for HTL_{1234} were 0.49, 0.44, and 0.41, respectively. For HTL_{346}, the largest Cohen's *d* values obtained by geometric mean kurtosis, mean kurtosis, and median kurtosis were 0.73, 0.72, and 0.65, respectively. For both HTL_{1234} and HTL_{346}, the geometric mean kurtosis performs better than the other two metrics, and the mean kurtosis performs better than the median kurtosis. The good performance of the geometric mean kurtosis is because most of the kurtosis sequences follow a lognormal distribution, and the geometric mean reflects the average level of data obeying a lognormal distribution, which makes it suitable for evaluating NIHL. The reason that the mean kurtosis is superior to the median kurtosis may be that the mean kurtosis can reflect the influence of extreme values in the sequence (strong impulsive components of noise exposure) on hearing, and the median kurtosis is likely to erase this effect.

On the other hand, Cohen's *d* value for HTL_{346} was much larger than that for HTL_{1234} (0.73 vs 0.49), which implies that the NIHL definition involving higher frequencies (3–6 kHz) is more sensitive for NIHL evaluation than the definition involving lower frequencies (1–4 kHz). Similar results can be obtained from the outcomes of algorithm 2.

C. Summary and prospects for future work

This study explored the correlation between the window duration for kurtosis computation and hearing loss evaluation and found that it is most appropriate to use a short window (between 40 s and 2 min) to compute kurtosis. It is suggested to take 60 s as the window length to calculate kurtosis and the geometric mean of kurtosis sequence as the kurtosis metric to evaluate NIHL. The results presented in this study have significance for guiding future research on occupational NIHL evaluation using kurtosis. It is worth noting that kurtosis is dependent on both the duration of the window over which the calculation is made and the sampling rate at which the noise waveform is recorded.

To highlight the impact of kurtosis values based on different time windows on the prediction model, the model features were simplified to include only a few features that are most relevant to NIHL, namely, age, exposure duration, and L_{Aeq}. These models might perform better if more features associated with hearing loss, such as the subjects' octave band noise levels, blood pressure, and smoking histories, were considered. If more data of individuals exposed to noise were collected in future work, a larger database and more diversified subjects could be involved in the study. Then the conclusions of this study could be further verified and refined.

ACKNOWLEDGMENTS

Y. T. and W. Q. designed and performed the experiments and wrote the paper; W. D. analyzed data and wrote the paper; M. Z. and T. Z. collected data and provided interpretive analysis; W. Q. and J. L. provided critical revision and the final approval of the version to be published. All authors discussed the results and implications and commented on the manuscript at all stages. We thank all reviewers and editors who helped to improve this work. This work was partially supported by Grant Nos. 200-2015-M-63857 and 200-2016-M-91922 from the United States National Institute for Occupational Safety and Health; Grant No. N00014-17-1-2198 from the United States Office of Naval Research; Grant No. 2015C03039 from the Key Research and Development Program of Zhejiang Province, China; Grant Nos. 81801796 and 81771936 from the National Natural Science Foundation, China; Grant No. 2020ND8AD01 from the Major Scientific Project of Zhejiang Lab, China; and Grant No. 2020QNA5031 from the Fundamental Research Funds for the Central Universities, China.

[1]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0003954 for detailed information on the methods and results.

Ahroon, W. A., Hamernik, R. P., and Davis, R. I. (**1993**). "Complex noise exposures—An energy analysis," J. Acoust. Soc. Am. **93**(2), 997–1006.

Centers for Disease Control and Prevention (**2019**). "Occupational hearing loss surveillance," https://www.cdc.gov/niosh/topics/ohl/ (Last viewed 25 June 2020).

Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Sciences* (Routledge, New York).

Davis, R. I., Qiu, W., and Hamernik, R. P. (**2009**). "Role of the kurtosis statistic in evaluating complex noise exposures for the protection of hearing," Ear. Hear. **30**(5), 628–634.

Davis, R. I., Qiu, W., Heyer, N. J., Zhao, Y., Yang, M. Q., Li, N., Tao, L., Zhu, L., Zeng, L., and Yao, D. (**2012**). "The use of the kurtosis metric in the evaluation of occupational hearing loss in workers in China: Implications for hearing risk assessment," Noise Health **14**(61), 330–342.

Dobie, R. A. (**1995**). "Prevention of noise-induced hearing loss," Arch. Otolaryngol. Head Neck Surg. **121**(4), 385–391.

Dobie, R. A. (**2001**). *Medical-Legal Evaluation of Hearing Loss*, 2nd ed. (Singular, San Diego, CA).

Goley, G. S., Song, W. J., and Kim, J. H. (**2011**). "Kurtosis corrected sound pressure level as a noise metric for risk assessment of occupational noises," J. Acoust. Soc. Am. **129**(3), 1475–1481.

Hamernik, R. P., Qiu, W., and Davis, B. (**2003**). "The effects of the amplitude distribution of equal energy exposures on noise-induced hearing loss: The kurtosis metric," J. Acoust. Soc. Am. **114**(1), 386–395.

ISO-1999 (**2013**). "Estimation of noise-induced hearing loss" (International Organization for Standardization, Geneva, Switzerland).

Johnson, M. E., and Lowe, V. W. (**1979**). "Bounds on the sample skewness and kurtosis," Technometrics **21**(3), 377–378.

Lei, S. F., Ahroon, W. A., and Hamernik, R. P. (**1994**). "The application of frequency and time-domain kurtosis to the assessment of hazardous noise exposures," J. Acoust. Soc. Am. **96**(3), 1435–1444.

National Institute for Occupational Safety and Health (**1998**). "Criteria for a recommended standard: Occupational noise exposure" (National Institute for Occupational Safety and Health, Cincinnati, Ohio).

Nelson, D. I., Concha-Barrientos, M., Driscoll, T., Steenland, K., Fingerhut, M., Punnett, L., Prüss-Üstün, A., Leigh, J., and Corvalan, C. (**2005**). "The global burden of selected occupational diseases and injury risks: Methodology and summary," Am. J. Ind. Med. **48**(6), 400–418.

Qiu, W., Hamernik, R. P., and Davis, B. (**2006a**). "The kurtosis metric as an adjunct to energy in the prediction of trauma from continuous, nonGaussian noise exposures," J. Acoust. Soc. Am. **120**(6), 3901–3906.

Qiu, W., Hamernik, R. P., and Davis, R. I. (**2013**). "The value of a kurtosis metric in estimating the hazard to hearing of complex industrial noise exposures," J. Acoust. Soc. Am. **133**(5), 2856–2866.

Qiu, W., Hamernik, R. P., Davis, B., and Zhao, Y. (**2006b**). "Research progress of kurtosis in the evaluation of noise-induced biological effects," Nat. Med. J. China **86**(11), 785–789.

Šidák, Z. (**1967**). "Rectangular confidence regions for the means of multivariate normal distributions," J. Am. Stat. Assoc. **62**(318), 626–633.

Smalt, C. J., Lacirignola, J., Davis, S. K., Calamia, P. T., and Collins, P. P. (**2017**). "Noise dosimetry for tactical environments," Hear. Res. **349**, 42–54.

Vapnik, V. N. (**2000**). *The Nature of Statistical Learning Theory* (Springer, New York).

Xie, H. W., Qiu, W., Heyer, N. J., Zhang, M. B., Zhang, P., Zhao, Y. M., and Hamernik, R. P. (**2016**). "The use of the kurtosis-adjusted cumulative noise exposure metric in evaluating the hearing loss risk for complex noise," Ear Hear. **37**(3), 312–323.

Zhang, M., Xie, H., Zhou, J., Sun, X., Hu, W., Zou, H., Zhou, L., Li, J., Zhang, M., Kardous, C. A., Morata, T. C., Murphy, W. J., Zhang, J. H., and Qiu, W. (**2020**). "New metrics needed in the evaluation of hearing hazard associated with industrial noise exposure," Ear Hear. **42**, 290–300.

Zhao, Y., Li, J., Zhang, M., Lu, Y., Xie, H., Tian, Y., and Qiu, W. (**2019a**). "Machine learning models for the hearing impairment prediction in workers exposed to complex industrial noise: A pilot study," Ear. Hear. **40**(3), 690–699.

Zhao, Y., Tian, Y., Zhang, M., Li, J., and Qiu, W. (**2019b**). "Development of an automatic classifier for the prediction of hearing impairment from industrial noise exposure," J. Acoust. Soc. Am. **145**(4), 2388–2400.

Zhao, Y. M., Qiu, W., Zeng, L., Chen, S. S., Cheng, X. R., Davis, R. I., and Hamernik, R. P. (**2010**). "Application of the kurtosis statistic to the evaluation of the risk of hearing loss in workers exposed to high-level complex noise," Ear Hear. **31**(4), 527–532.