4-2017

How Tonality and Loudness of Noise relate to Annoyance and Task Performance” Noise Control Eng. J. 65(2), 71-82.

Joonhee Lee
University of Nebraska - Lincoln, gi20plus@gmail.com

Jennifer M. Francis
University of Nebraska - Lincoln, jmfrancis624@gmail.com

Lily M. Wang
University of Nebraska - Lincoln, lwang4@unl.edu

How tonality and loudness of noise relate to annoyance and task performance

Joonhee Lee[a], Jennifer M. Francis[b] and Lily M. Wang[b]

(Received: 4 September 2015; Revised: 9 November 2016; Accepted: 7 February 2017)

Audible tones in noise generated by building mechanical equipment can be a leading cause of complaints from occupants. A number of metrics have been developed to quantify prominence of a tone, but previous work has shown that the impact of a certain tonality appears to vary with the level of the broadband noise signal. More work on how tonal signals of varying tonality, tone frequency and broadband noise levels relate to annoyance and task performance is needed. This paper investigates such relationships between current noise metrics, annoyance and task performance under assorted tonal noise conditions through subjective testing. Participants rated their perceived annoyance after being exposed to noise signals with differing levels of tones while solving Sudoku puzzles. In addition to assessing annoyance, the test also surveyed the perceived workload caused by the noise by using a modified noise-induced task load index questionnaire. Five levels of tonal prominence for each of two tonal frequencies were added above two different ambient background noise levels to create 20 noise signals of interest. The task performance results based on the Sudoku puzzle answers show trends of decreasing accuracy with increasing tone strengths, but the differences are not statistically significant. Other findings are that loudness metrics are most highly correlated with annoyance responses, while tonality metrics demonstrate relatively less but also significant correlation with annoyance. Generally, participants felt more annoyed with higher background noise levels, lower tone frequency and more prominent tone strength. Based on correlation analysis, a multiple regression model using two of the most strongly correlated noise metrics, ANSI loudness level and tonal audibility, has been developed for predicting annoyance responses from tonal noise conditions

Primary subject classification: 63.2; Secondary subject classification: 13.1

1 INTRODUCTION

Most mechanical systems in buildings generate significant tones due to rotating components. HVAC (heating, ventilating and air conditioning) equipment in buildings are becoming more energy-efficient, but these changes are often accompanied with changing sound quality including more prominent tones. Increasing the tonality of the noise, though, can result in increased complaints from building occupants and neighbors, but quantitative data published to date are not able to establish evidence-based guidelines or limits for tones in different levels of building equipment noise. Noise regulations in many municipalities in the United States apply a 5 dB penalty if tones are detected using a one-third octave band measurement technique given in ISO 1996-2:2007 Annex D[1], when comparing against maximum allowed noise levels[2–5]. However, the one-third octave band measurement technique is not always capable of detecting a tonal component, if the tone falls on the edge of two bands. The 5 dBA penalty value is also rather arbitrary as that value has not been determined from psychoacoustic studies; the same 5 dB penalty is applied once a tone is deemed to be prominent, but more prominent tones are not penalized more greatly than less prominent ones.

A considerable amount of literature has been published on the relationship between tones in noise and human annoyance, as perceptible tones in noise from aircraft, office equipment and wind turbines have been recognized as

a) Charles W. Durham School of Architectural Engineering and Construction, University of Nebraska — Lincoln, Omaha, Nebraska, 68182, USA. Department of Building, Civil and Environment Engineering, Concordia University, Quebec, Montreal, H3G 2W1 CANADA; email: joonhee.lee@concordia.ca.

b) Charles W. Durham School of Architectural Engineering and Construction, University of Nebraska — Lincoln, Omaha, Nebraska, 68182, USA.

serious sources of public noise pollution since the 1960s. In the 1980s, Hellman found that tonal components in broadband spectra impact ratings of annoyance, loudness and noisiness[6] and that the number of tones and frequency differences between tones as well as the frequency of the tone itself influence annoyance[7]. More and Davies[8] also examined the effects of tones on human annoyance from aircraft flyover noise, using a questionnaire without any accompanying tasks. They found that regression models that include metrics for both loudness and tonality matched well with annoyance responses from their time-varying signals. Lee et al.[9] investigated the tonality perception for harmonic complex tones and pointed out the difficulties of quantifying overall tonality including harmonic tones with existing methods. Hastings et al.[10] investigated assorted tonality metrics for predicting tonality and annoyance of noises. They proposed modifications in calculating the existing metrics and suggested that the bandwidth and roll-off rate of tones should be included for accurate tonality perception for aircraft noise.

More recent attention has focused on perception of tones in noise from building machinery. Ryherd and Wang[11] investigated assorted building mechanical noise samples and showed that current indoor noise criteria were not accurately reflecting annoyance because the criteria do not typically account for tonal characteristics in assessment. Susini et al.[12] used multidimensional scaling analysis to find that one of the most important sound quality dimensions of noise from indoor air-conditioning units is the ratio of tonal harmonic components to broadband noise components. Berglund et al.[13] also investigated perception of environmental noises including ventilation-like noise spectra with the multidimensional scaling methodology and concluded that spectral contrast, which is related to the tonality, is the best acoustic index for predicting the preference rating of noises. Besides laboratory studies, Landström et al.[14] explored noise levels and annoyance by occupants in actual working spaces. They found that the relation between noise levels and annoyance was weak, but annoyance ratings were significantly increased when tones were present in the noise. These previous studies strongly suggest that tonality metrics should be included when evaluating noise from mechanical systems in buildings, but to date none of existing tonal metrics is utilized broadly and there is still limited understanding in linking measurable objective metrics to annoyance.

Besides annoyance, how tones in noise impact task performance is also of interest. Previous research findings into effects of tones on human performance have been inconsistent and limited. Landström et al.[14,15] found that task performance was significantly lower for tonal noises and Laird[16] argued that tones above 512 Hz have a greater effect on increasing error rates of tasks in artificial factory experiments. Grjmaldi[17] also found tendencies of slower response times and increasing error rates of coordinated movement performance for tones in the range of 2400 to 4800 Hz. However, a few other studies[11,18] did not find any statistically significant differences in task performance between broadband and tonal noises.

This paper describes a subjective investigation on how exposure to tonal noise as produced by building mechanical systems impacts human annoyance and task performance, using a larger variety of signals than most previous studies. The relationships between a number of known noise metrics, objectively describing both loudness and tonality and annoyance responses, are examined. Results are also used to develop a preliminary annoyance prediction model through statistical analysis, based on a noise signal's loudness and tonality. While harmonic structures of tones have been shown to impact annoyance and other psychoacoustic qualities such as sharpness, roughness or fluctuation may play a part as well, those aspects were not directly considered in this investigation. Rather, this study focused on how these two primary characteristics of loudness and tonality affect annoyance and performance because previous studies pointed out that the tonality, impulsivity and loudness have the most influential impacts on listeners' responses.

There is a degree of uncertainty in defining annoyance due to noise. ISO/TS 15666:2003 defines noise-induced annoyance as “one person's individual adverse reaction to noise in various ways including dissatisfaction, bother, annoyance and disturbance”[19]. While a variety of definitions for annoyance have been suggested, it is generally agreed that annoyance is concerned with physical noise characteristics, the context of measurement and personal attributes of listeners[20]. In this study, the physical noise characteristics of interest are loudness and tonality. Although the subjective testing has been conducted in a controlled laboratory, the context of the measurement is meant to be like an office environment. From reviewing previous research studies, Marquis-Favre et al.[21] indicated that, among non-acoustic factors that can influence annoyance, fear and noise sensitivity were found to have the most significant effects. In the investigation discussed herein, fear was not considered since listeners are not expected to fear regular levels of building mechanical noise, but noise sensitivity was surveyed as a personal attribute.

The noise metrics investigated in this paper that have been developed to quantify tonality or the degree to which tones are present in broadband noise are reviewed. ANSI S12.10-2010/Part1 Annex D[22] presents tone-to-noise ratio (TNR) and prominence ratio (PR) to quantify tonality and ISO 1996-2:2007 Annex C[5] suggests tonal audibility (ΔL_{ta}). These metrics are calculated from the steady-state frequency spectrum of the noise recording through digital fast Fourier transform analysis. There are two main differences between tonal audibility and the previous two metrics, tone-to-noise ratio and prominence ratio. One major difference is that tonal audibility uses A-weighted sound pressure levels and includes a frequency correction term in its calculation so that the prominence criteria of tones is constant across frequencies, whereas TNR and PR ratings are based on unweighted sound pressure levels. Consequently, the prominence of tones is frequency dependent for TNR and PR ratings, but not for ΔL_{ta}. That is, PR = 5 for a 100 Hz tone is not necessarily the same perceived tonality as PR = 5 for a 500 Hz tone. The other difference is that the tonal audibility uses a linear regression line instead of actual noise components when calculating masking tonal levels within the critical bands. The equation to calculate ΔL_{ta} is given by:

$$\Delta L_{ta} = L_{pt} - L_{pn} + 2\ \text{dB} + \log\left[1 + \left(\frac{f_c}{502}\right)^{2.5}\right], \quad (1)$$

where L_{pt} is the total sound pressure level of the tones; L_{pn} is the total sound pressure level of the masking noise in the critical band; and f_c is the center frequency of the critical band. Based on the tonal audibility calculation, penalty factors between 0 and 6 dB are provided to adjust the overall A-weighted noise levels, rather than setting prominence criteria. It also requires separate analysis for each tone within a multi-tonal noise signal. Aures' tonality (Aures) is another metric for tonality that considers the frequency, as well as bandwidth and levels of all tonal components through use of weighting functions[23]. It is one of the few that can account for multiple tones in a signal.

Popular loudness metrics are also investigated in this study because previous studies have found that loudness of the noise is the most relevant feature correlating to annoyance besides tonality. Among the included loudness metrics are A-weighted (dBA) and unweighted (dB) equivalent sound pressure levels and stationary loudness levels calculated according to ANSI S3.4-2007[24] (ANSI loudness) and ISO 532-1975 B method[25] (ISO loudness). The ISO loudness and ANSI loudness are based on Zwicker's[26] and Glasberg and Moore's[27] loudness models respectively. They both can use stationary one-third octave band data for the calculation of loudness.

A few noise metrics that consider both loudness and tonality to produce an overall rating for tonal noises have been proposed. These combined metrics basically add penalty values to the loudness levels due to the presence of tones. The Joint Nordic Method (JNM) is standardized in ISO 1996-2:2007[1], where the penalty k values are derived from tonal audibility and added to A-weighted sound pressure level. Perceived noise level (PNL) was implemented to quantify subjective annoyance of aircraft noise, calculated from one-third octave band values; tone-corrected perceived noise level (PNLT) is a revised version of PNL with the addition of a tone correction factor[28]. Sound quality indicator (SQI) is a similar metric suggested by the Air-Conditioning, Heating and Refrigeration Institute to rate the sound quality of building mechanical product noise based on one-third octave bands[29], but it has yet to be applied widely.

2 METHODOLOGY

2.1 Test Laboratory

The subjective testing was completed in an acoustic testing chamber at the University of Nebraska. Figure 1 illustrates a schematic plan of the testing chamber, which has a volume of approximately 27.8 m^3. The chamber is acoustically isolated from a monitor room and nearby spaces. Materials in the room include carpeted floor, gypsum board walls with additional absorptive panels, acoustic bass traps and acoustical ceiling tiles. The average mid-frequency reverberation time is 0.31 seconds and the ambient background noise level is 37 dBA when air-conditioning in the chamber is turned off. Figure 2 presents the ambient background noise levels in the chamber across octave bands. The tonal test signals were generated through a ceiling-mounted Armstrong i-ceiling speaker and a sub-woofer in a corner. The i-ceiling speaker appears as other ceiling tiles in the ceiling grid, so that participants cannot visually identify the location of the sound source. Participants sat in the middle of the chamber and were advised not to move their location during the experiment.

2.2 Test Signals

A total of 22 noise signals were generated for use in this study by the program Test Tone Generator from Esser Audio. Two levels of broadband noise without any

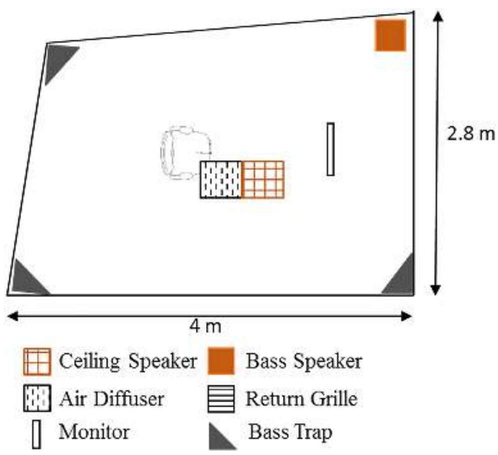

Fig. 1—Schematic plan of the acoustic testing chamber at the University of Nebraska.

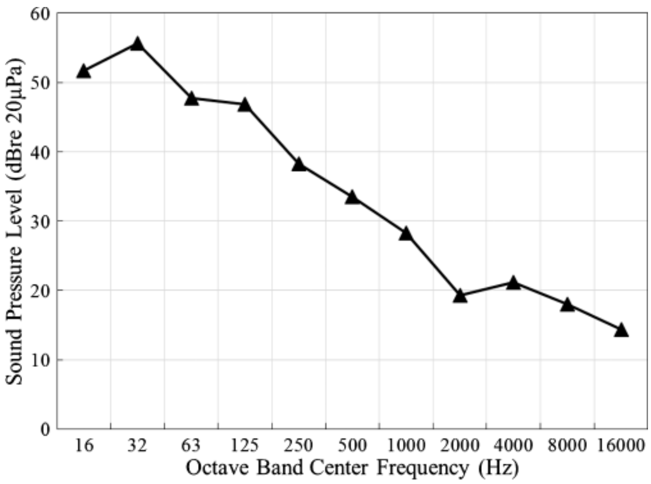

Fig. 2—Measured octave band spectra for the ambient background noise in the test chamber when air-conditioning is off.

tonal components were used: either 40 or 55 dBA overall, following a −5 dB/octave room criteria (RC) contour[30]. These two levels are in the range of common background noise levels found in buildings. A single tone at one of two frequencies and at one of five prominence levels was added separately to the broadband noise signals, to create the other 20 noise signals. The two tonal frequencies were selected to be 125 Hz, which is a common tone generated by building mechanical equipment and 500 Hz as it is slightly higher but still in the frequency range where a number of other building mechanical equipment exhibit tones. The five tone levels were selected to range from below to above the prominence thresholds listed in ANSI S12.10-2010[22]: PR = 18 dB for 125 Hz and PR = 12 dB for 500 Hz. Table 1 presents the prominence ratio values for each test signal. Figure 3 illustrates the one-third octave band spectra of the test signals. All tonal signals were measured using a

Table 1—Prominence ratios for the tones in the noise stimuli used in the subjective testing as listed by tonal frequency, broadband background noise level and tone level.

Frequency (Hz)	BNL (dBA)	Prominence ratio (dB)				
		Tone level 1	Tone level 2	Tone level 3	Tone level 4	Tone level 5
125	40	15	18	21	24	27
	55	13	15	18	21	24
500	40	9	12	15	18	21
	55	6	9	12	15	18

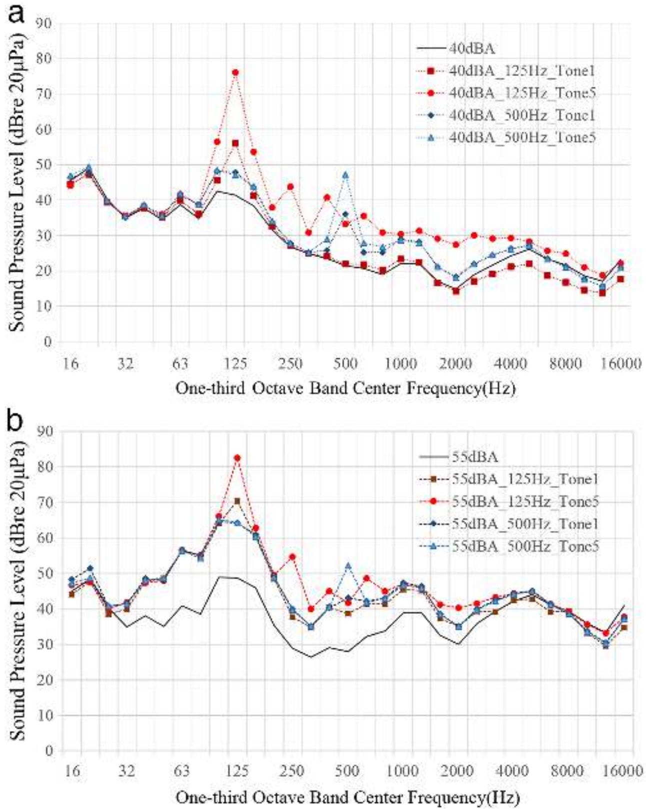

Fig. 3—Measured one-third octave band spectra for a few of the test noise signals: (a) Broadband 40 dBA signal and some with assorted tones; (b) Broadband 55 dBA signal and some with assorted tones. Tones were either at 125 or 500 Hz; for clarity, only the lowest and highest tonal strengths are presented.

B&K 4189-A microphone through the B&K PULSE system at the listener's ear position in the testing chamber and averaged over a minute for calculation of noise metrics. The metrics were calculated using Matlab or programs provided by the associated standards.

2.3 Test Participants and Procedure

Ten participants, four females and six males, were recruited from the University of Nebraska — Omaha community, ranging in age from 25 to 43 years old. The University of Nebraska — Lincoln Institutional Review Board approved the study and each participant was paid for their time. The sample size was determined by a priori power analysis using the effect size from More and Davies'[8] statistical results using G*Power version 3.1[31]. The effect size for multiple regression models, Cohen's f^2, was calculated as 6.69 from the squared multiple correlation values in the previous study. The minimum sample size was then found to be six participants to achieve 80% power ($1 - \beta$) at two-sided 5% significance level (α). Based on this finding and available research funds, a testing plan was designed to assess 22 signals across ten test subjects.

All participants completed an orientation session including a hearing screening test before participation and demonstrated normal hearing with thresholds below 25 dB hearing level (HL) from 125 Hz to 8 kHz. The noise sensitivity of each participant was also measured by a reduced version (13 items only) of the Noise-Sensitivity-Questionnaire (NoiSeQ) by Schutte et al.[32] during the orientation session. The participants were asked to answer each item using a four-level rating scale (strongly agree = 1, slightly agree = 2, slightly disagree = 3, strongly disagree = 4). The responses were averaged across all items to form a composite scale to quantify the noise sensitivity for each participant.

The main test consisted of two parts: a direct assessment with task (part A) and a magnitude adjustment test (part B). The results of part B have been presented in another paper[33] and hence are not included herein. In part A, participants were asked to complete as many Sudoku number puzzles as possible while exposed to a broadband noise signal, some with assorted tonal components, for 10 minutes. Sudoku puzzles were selected as the measure of task performance, as they are compact to administer, easy to explain to test participants and have been used as a measure of task performance in other studies with results showing significant relationship with working memory[34,35]. All participants practiced solving Sudoku puzzles during the orientation session before participating in the main test and the difficulty of all Sudoku puzzles in the main test was held constant. The puzzles were all nine by nine with forty of the eighty-one grids being prefilled with numbers.

After spending 10 minutes solving the Sudoku puzzles, the subjects answered five questions on a subjective questionnaire about the noise they had just heard. The questionnaire was a modified version of the NASA task load index[36]. The original NASA task load index is divided into six subscales: mental demand, physical demand, temporal demand, performance, effort and frustration. In this study, the questions on physical demand, temporal demand and frustration were not included; instead questions were added on rating loudness and annoyance incurred by noise as shown in Table 2. Participants responded to each question based on a 21-point scale (to match the scale from the original NASA task load index) on a paper form.

Part A consisted of ten 30-minute sessions that were completed by each subject individually on different days. Within each 30-minute session, subjects were exposed to three noise signals (each for 10 minutes) and thus completed three sequences of Sudoku puzzles (different puzzles each time) followed by the questionnaire. To minimize the influence of back-to-back comparisons of tonal noise conditions, a neutral background noise condition without any tonal components was used as the second signal within each 30-minute test session. Within a single 30-minute test session, the noise level of the broadband noise without consideration for any tonal components remained at a constant level, either 40 or 55 dBA. The presentation order of the background noise levels and tonal test signals was carefully balanced across all subjects using a Latin square design.

Table 2—Items from the subjective questionnaire, as modified from the NASA task load index.

Description	Questions
Mental demand	1. How mentally demanding was the task?
Overall performance	2. How successful were you in accomplishing what you were asked to do?
Effort	3. How hard did you have to work to accomplish your level of performance?
Loudness	4. How loud was the noise?
Annoyance	5. How annoying was the noise?

Two task performance measures were gathered by (1) counting the amount of Sudoku puzzles a subject completed within a 10-minute trial, with partial completions included as well, and (2) quantifying the accuracy of the puzzle answers in terms of correct numbers among those answered in a puzzle. The maximum and minimum number of Sudoku puzzles participants completed in one 10-minute session were 2.8 and 0.3, and the maximum and minimum accuracy of the puzzles were 100% and 69%.

3 RESULTS AND DISCUSSION

The reliability of each participant's responses was determined from correlation analysis of the participant's individual annoyance responses to a loudness metric, tonality metric and average ratings across participants[8]. Figure 4 presents correlation coefficients of each participant's annoyance responses to the ANSI loudness level, tonal audibility and mean values across participants. Two participants' responses (number 6 and 8) were excluded from all analyses because they rated responses randomly regardless of sound characteristics (correlation value <0.2). The subject-to-loudness coefficient of participant 6 was 0.17 and the subject-to-

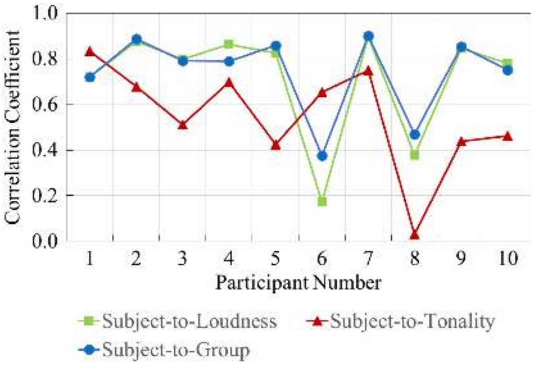

Fig. 4—Correlation coefficients of each participant's annoyance responses to each signal's ANSI loudness, tonal audibility and group average.

tonality correlation coefficient of participant 8 was 0.03. All subsequent analyses are based on the remaining eight subjects, which is still above the minimum sample size of six recommended from the a priori power analysis. Additionally, outlier responses of annoyance and task performance scores were excluded from the statistical analysis presented below. The outliers were identified using the criterion of being beyond three standard deviations from the average across participants. Two outliers of annoyance ratings and three outliers of task performance scores were excluded for analyses based on this criterion.

3.1 Relating Task Performance to Subjective Responses and to Noise Attributes

The task performance measures related to the Sudoku puzzles were correlated to the participants' subjective responses on the modified NASA task load index questionnaires (Table 3). Spearman's correlation (ρ) was utilized because not all of the variables met the assumption of having a normal distribution with the sample size utilized. An additional "TLX-avg" score was calculated as the averaged value of all five items from the modified survey to represent an overall rating of subjective task load perception induced by noise exposure. Since the task difficulty was held constant with equivalently difficult Sudoku puzzles throughout the experiment, the variations in subjective ratings observed within subjects can be considered as the result of varying background noise conditions. Job et al.[37] have recommended against using a single question item about annoyance because of its reduced validity; consequently, the composite modified Noise TLX rating is proposed as an alternative in this laboratory study. With a Cronbach's α coefficient for the reliability of 0.82 and a test–retest correlation of the Noise TLX measure for the stability of 0.77, the "TLX-avg" questionnaire was found to be internally consistent and stable over time and thus suitable for the purpose of this test.

As Table 3 indicates, most of the subjective responses were significantly correlated with each other. Specifically of interest, the mental demand responses showed high correlations with perceptions of loudness and annoyance of the noise, and as expected, loudness and annoyance ratings were significantly correlated with each other (ρ = 0.948). The only statistically significant correlation between a task performance result and a subjective response was between "accuracy" (accuracy rates of participants' puzzle answers) and responses to the "performance" question on the questionnaire (ρ = −0.483). Figures 5 and 6 present the averaged task performance of the accuracy and number of completed puzzles

Table 3—Spearman's correlation analysis of the subjective responses and Sudoku puzzle task performance measures. TLX-avg is the average value of the responses to all five questions on the modified task load index questionnaire. "No. of completed" refers to the number of completed puzzles for each trial and "accuracy" indicates accuracy rates of participants' puzzle answers.

	Mental demand	Performance	Effort	Loudness	Annoyance	TLX-avg	No. of completed	Accuracy
Mental demand	–							
Performance	0.260	–						
Effort	0.610*	0.496**	–					
Loudness	0.501**	0.105	0.230	–				
Annoyance	0.528**	0.162	0.398	0.948*	–			
TLX–avg	0.631*	0.374	0.601*	0.880*	0.956*	–		
No. of puzzles completed	−0.317	−0.438	−0.394	0.074	−0.020	−0.171	–	
Accuracy	−0.105	−0.483**	−0.071	−0.289	−0.252	−0.330	0.080	–

*Correlation is significant at the 0.01 level (2-tailed).
**Correlation is significant at the 0.05 level (2-tailed).

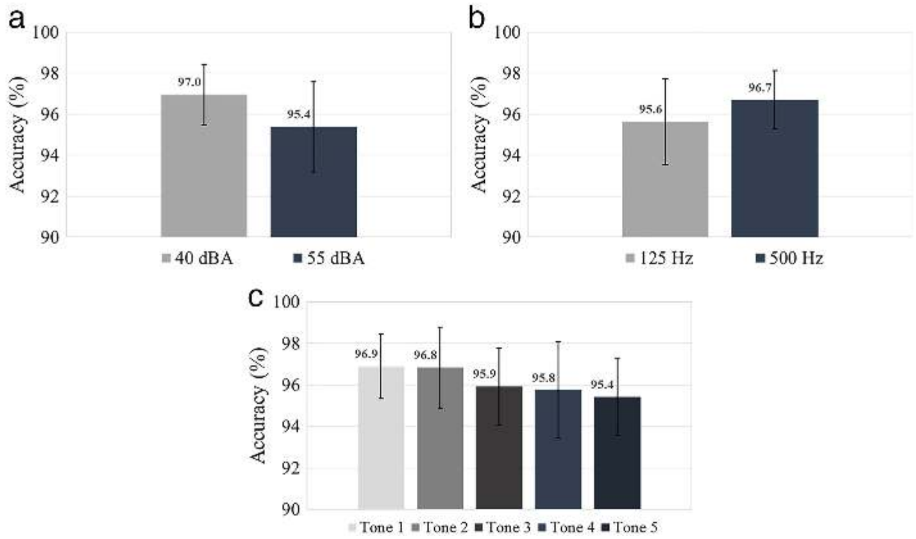

Fig. 5—Averaged accuracy of Sudoku puzzle answers as task performance scores plotted against (a) Background noise level, (b) Tonal frequency and (c) Strength of the tones, where Tone 1 indicates the least prominent tone and Tone 5 indicates the most prominent tone. Error bars indicate one standard error.

against the physical attributes of the noise signals, including background noise level, tone frequency and the five levels of tone strengths. The repeated measure ANOVA (analysis of variance) confirms that there were no statistically significant differences between task performances across the various noise attributes. Thus, subjects did not complete more puzzles or have higher accuracy under any particular tonal frequency, background noise level or tone strength, although there appears to be a slight tendency of lower accuracy with greater tone strength.

3.2 Relating Noise Attributes to Annoyance Responses

To understand how the physical aspects of the noise signals (background noise level, tone frequency and tonal strength) related to annoyance, a three-way repeated

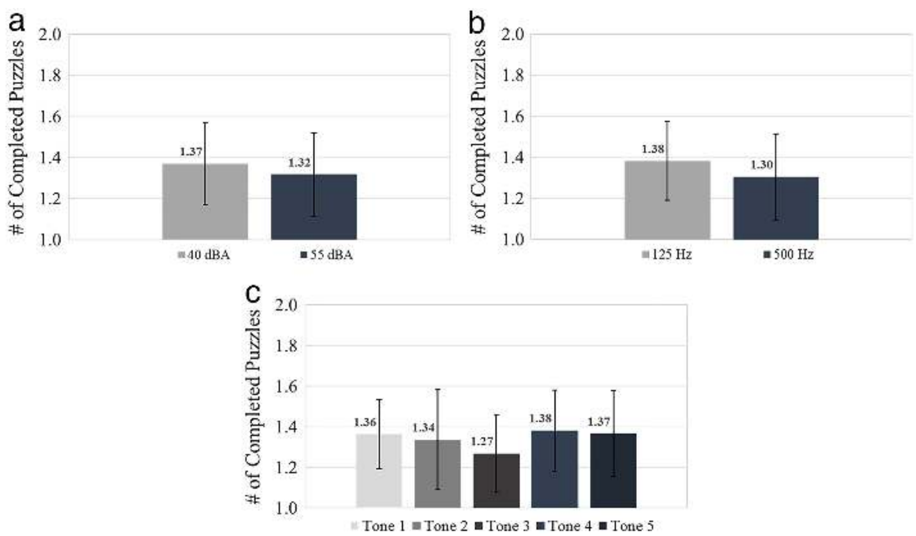

Fig. 6—Averaged number of completed Sudoku puzzles as task performance scores plotted against (a) Background noise level, (b) Tonal frequency and (c) Strength of the tones, where Tone 1 indicates the least prominent tone and Tone 5 indicates the most prominent tone. Error bars indicate one standard error.

measure ANOVA was conducted. Mauchly's test indicated that the assumption of sphericity had been met for the main effects of tonal strength and its interactions with tone frequency and background noise level. The analysis shows a significant main effect of background noise level [$F(1,7) = 82.61$, $p < 0.001$, $\eta_p^2 = 0.92$], tone frequency [$F(1,7) = 20.01$, $p = 0.003$, $\eta_p^2 = 0.74$] and tonal strength [$F(4,28) = 4.76$, $p = 0.005$, $\eta_p^2 = 0.41$] on annoyance.

The main analysis shows that the 55 dBA based tonal signals were significantly more annoying than 40 dBA based tonal signals and that the 125 Hz tonal signals were significantly more annoying than 500 Hz tonal signals. Contrast comparisons reveal that the 4th highest [$F(1,7) = 10.420$, $p = 0.014$] and 5th highest [$F(1,7) = 12.069$, $p = 0.010$] in prominence tonal signals were perceived as more annoying than the least (1st) prominent tonal signals.

Figure 7 illustrates the mean annoyance ratings across background noise levels, tonal frequencies and tone strengths. Summarizing these results, the overall background noise level does impact annoyance, with higher levels leading to greater annoyance. The lower frequency tone generated greater annoyance ratings, but one should note that the prominence levels of the 125 Hz tone versus those of the 500 Hz tone used in the study were not the same even though the relative differences from the threshold of tones presented in ISO 1996-2:2007 are the same. There was also a significant interaction effect between background noise level and tone frequency [$F(1,7) = 33.31$, $p = 0.014$, $\eta_p^2 = 0.60$]. As plotted in Fig. 7(d), the difference between annoyance ratings of the 125 and 500 Hz tones was greater with the 40 dBA background noise level condition than with the 55 dBA background noise level condition. It appears that tonal frequency is less related to annoyance at higher background noise levels, but plays a larger role at lower background noise levels.

The data on tonal strength shows that higher tone levels are linked to higher annoyance ratings; analysis of the data to determine a threshold of annoyance is presented in Francis et al.'s study[33]. Noise sensitivity was expected to be associated with annoyance but did not demonstrate statistically significant effects in the ANOVA analysis as a between-subjects factor. This is attributed to the limited number of subjects in the study, which was selected based on a power analysis of previous annoyance results, rather than noise sensitivity results.

3.3 Correlations of Noise Metrics with Subjective Responses

The previous section showed that physical aspects of the noise signals (specifically loudness and tonality) were correlated with annoyance responses; in this section, assorted metrics for quantifying those physical aspects are tested against the subjective responses. Spearman's nonparametric correlation coefficients were calculated between a number of noise metrics and the average participants' perception ratings of loudness, annoyance and TLX-avg. The results have been analyzed in two ways:

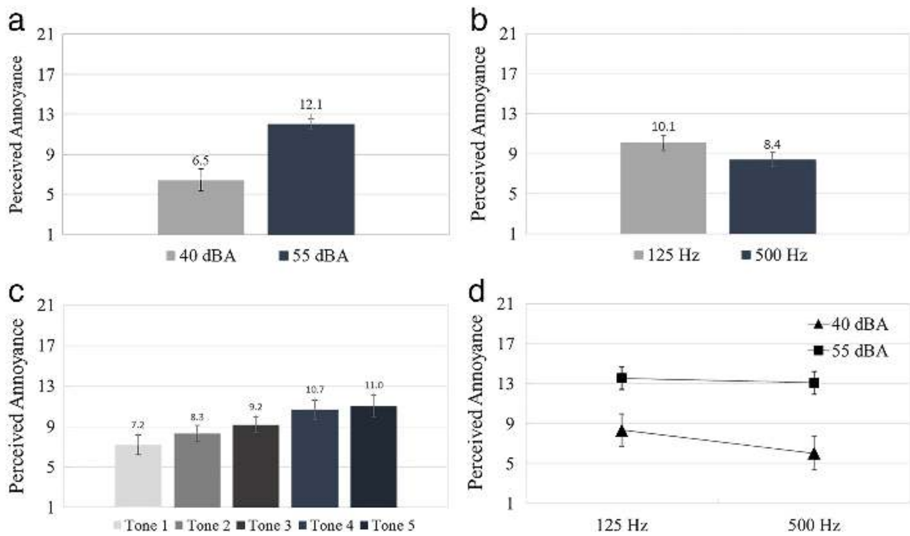

Fig. 7—Mean annoyance perception ratings plotted against (a) Background noise level, (b) Tonal frequency, (c) Strength of the tones, where Tone 1 indicates the least prominent tone and Tone 5 indicates the most prominent tone and (d) Interaction of background noise level and tonal frequency. Error bars indicate one standard error.

Table 4—Spearman's correlation analysis of noise metrics against subjective responses and Sudoku puzzle task performance. The results are analyzed first with all signals included and then in two groups separated by background noise level (40 or 55 dBA). Bolded values indicate metrics chosen for use in the regression model, based on their overall high significant correlation values.

All signals (40 dBA and 55 dBA BNL)			
	Loudness	Annoyance	TLX-avg
PR	0.150	0.186	0.147
TNR	−0.123	−0.081	−0.095
ΔL_{ta}	0.006	0.056	0.019
Aures	0.297	0.359	0.314
dB	0.805*	0.824*	0.772*
dBA	0.866*	0.887*	0.842*
ANSI loudness	**0.946***	**0.950***	**0.926***
ISO loudness	0.938*	0.952*	0.925*
PNL	0.892*	0.920*	0.886*
PNLT	0.869*	0.877*	0.826*
JNM	0.840*	0.869*	0.818*
SQI	0.904*	0.899*	0.856*
40 dBA BNL only			
PR	0.794*	0.867*	0.782*
TNR	0.794*	0.867*	0.782*
$\boldsymbol{\Delta L_{ta}}$	**0.778***	**0.888***	**0.815***
Aures	0.673**	0.709**	0.697**
dB	0.806*	0.939*	0.855*
dBA	0.794*	0.927*	0.830*
ANSI loudness	0.685**	0.745**	0.697**
ISO loudness	0.685**	0.745**	0.697**
PNL	0.685**	0.842*	0.867*
PNLT	0.794*	0.830*	0.758**
JNM	0.794*	0.927*	0.830*
SQI	0.806*	0.806*	0.709**
55 dBA BNL only			
PR	0.799*	0.867*	0.758**
TNR	0.709**	0.845*	0.845*
$\boldsymbol{\Delta L_{ta}}$	**0.787***	**0.891***	**0.818***
Aures	0.781*	0.903*	0.782*
dB	0.715**	0.756**	0.530
dBA	0.707**	0.770*	0.564
ANSI loudness	0.878*	0.855*	0.709**
ISO loudness	0.817*	0.867*	0.697**
PNL	0.720**	0.806*	0.539
PNLT	0.744**	0.782*	0.527
JNM	0.707**	0.770*	0.564
SQI	0.689**	0.663**	0.444

*Correlation is significant at the 0.01 level (2-tailed).
**Correlation is significant at the 0.05 level (2-tailed).

first with all twenty tonal signals included and then with the average ratings for ten signals grouped separately by the broadband background noise level (40 or 55 dBA). Table 4 presents correlation coefficients between all noise metrics with the subjective perception responses.

When analyzing all signals, the noise metric that demonstrates the highest correlation coefficients with the perceived loudness, annoyance and TLX-avg ratings is ANSI loudness level. Other loudness metrics were also significantly correlated to the perception ratings, but the tonality metrics such as prominence ratio, tone-to-noise ratio, tonality audibility and Aures' tonality did not statistically correlate or had lower coefficients than loudness metrics. This confirms that loudness is the most dominant factor in determining subjective perception of noise.

When the signals are grouped separately by broadband background noise levels, though, tonality metrics did show higher correlations with subjective ratings than loudness metrics. The coefficient values for the assorted tonality metrics are all very similar with no particular metric clearly performing better than others. However, when only looking for correlation coefficients with annoyance, tonal audibility showed slightly higher correlation coefficients than other tonality metrics (0.888 for 40 dBA BNL and 0.891 for 55 dBA BNL). Aures' tonality also showed high correlation with annoyance from 55 dBA BNL signals (0.903), but it showed lower correlation than other metrics with 40 dBA BNL signals (0.709). The results indicate that, when the broadband background noise level is controlled or comparable, tonality becomes a more influencing factor on annoyance evaluation. Figure 8 presents scatterplots of the averaged annoyance responses (a) with the ANSI loudness level across the entire group and (b) with tonal audibility, separated by background noise level.

For all cases, combined metrics such as the Joint Nordic Method, tone-corrected perceived noise level and sound quality indicator did not show remarkably better performance than loudness metrics, even though these combined metrics were significantly related with annoyance ratings. The results suggest that imposing penalty values to loudness levels based on tonal strength may not be the most appropriate way to quantify overall subjective annoyance of tonal noise. Instead, using separate metrics to account for tonality and loudness of building mechanical noises is an effective way to relate to the signal's annoyance.

3.4 Regression Model between Noise Metrics and Annoyance

Based on the results in Table 4, ANSI loudness level and tonal audibility were selected to be used as predictors

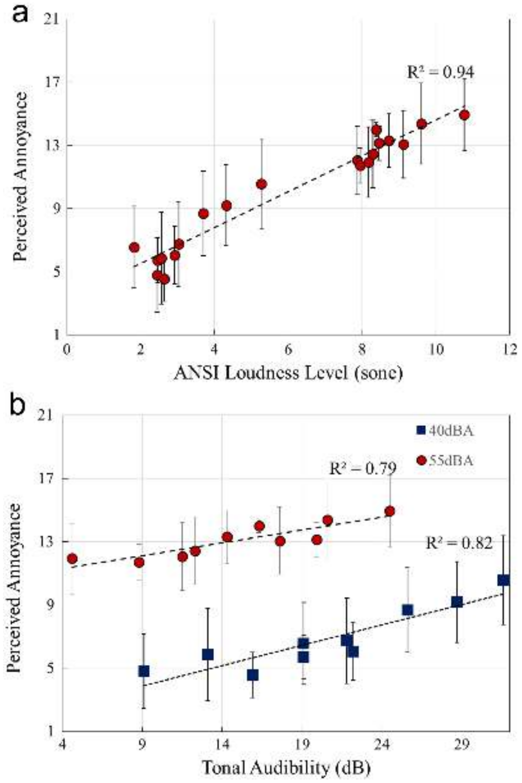

Fig. 8—Average (mark) and standard deviation (error bar) of the annoyance ratings for noise signals against (a) ANSI loudness level for all signals and (b) Tonal audibility for 40 and 55 dBA BNL separately. Dashed lines indicate regression lines of annoyance rating prediction with regard to each metric.

for a linear multiple regression model for annoyance, because these two metrics resulted in among the strongest correlations with annoyance perception compared to other noise metrics. Equation (2) presents the multivariate regression model with ANSI loudness level and tonal audibility.

$$\begin{aligned}\text{Annoyance} &= 1.806 + 1.164 \\ &\quad * [\text{ANSI Loudness (sone)}] \\ &\quad + 0.072 * [\text{Tonal Audibility (dB)}].\end{aligned} \tag{2}$$

Table 5 also presents standard error of coefficients, standardized coefficients and statistical significance when ANSI loudness level was only used (in step 1) and when tonal audibility was also included (in step 2), in addition to the coefficient values for each predictor. Standardized β values indicate the number of standard deviations that the outcome annoyance will change as a result of one standard deviation change in the predictor. The R^2 value for the first step model is 0.943, which is a measure of goodness-of-fit of linear regression, indicating that 94.3% of the annoyance rating variance can be explained by the ANSI loudness model only. When including tonal audibility as a second predictor, the R^2 value increased to 0.962. Even though this increase is small, the multivariate regression model does significantly predict more variation in annoyance perception when including tonal audibility as a second predictor; for step 2, the ANSI loudness level [$t(17) = 20.796$, $p < 0.001$] and tonal audibility [$t(17) = 2.943$, $p = 0.009$] are both significant predictors of annoyance. Figure 9 illustrates a regression line with the calculated linear model.

The results of the correlation analysis and regression model presented in this paper are in line with the findings from More and Davies' study[8], which focused on aircraft flyover noise rather than building mechanical system noise. Their work focused only on annoyance and used metric values that were exceeded some percentage (often 5%) of the time since their flyover

Table 5—Linear regression model of predictors for annoyance perception, with 95% bias corrected and accelerated confidence intervals reported in parentheses. Confidence intervals and standard errors are based on 1000 bootstrap samples. Standardized β values indicate the number of standard deviations that the outcome annoyance will change as a result of one standard deviation change in the predictor.

	b	Standard error B	*β*	*p*
		Step 1		
Constant	3.254(2.305, 4.310)	0.512		$p = 0.001$
ANSI loudness (sone)	1.137(1.004, 1.263)	0.066	0.971	$p = 0.001$
		Step 2		
Constant	1.806(0.498, 3.187)	0.683		$p = 0.020$
ANSI loudness (sone)	1.164(1.043, 1.308)	0.069	0.994	$p = 0.001$
Tonal audibility (dB)	0.072(0.027, 0.111)	0.021	0.141	$p = 0.004$

Note: 0.943 for Step 1; $\Delta R^2 = 0.019$ for Step 2.

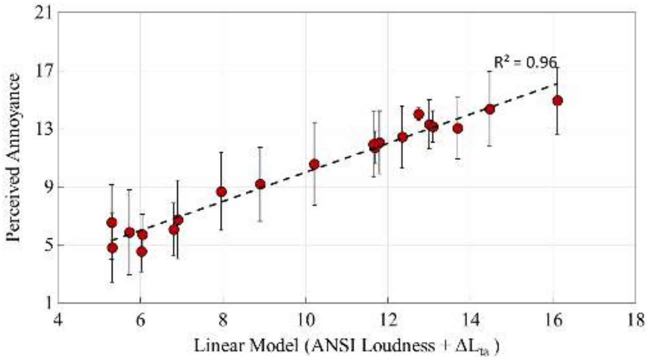

Fig. 9—Average (mark) and standard deviation (error bar) of the annoyance ratings for noise signals plotted against the proposed linear regression model of annoyance perception (dashed) based on ANSI loudness level and tonal audibility (R^2 = 0.96).

signals varied in time. They demonstrated that the best regression model, when using existing metrics to match their annoyance responses, included both a loudness metric and a tonality metric and they proposed revision to the penalty values used for the Joint Nordic Method to improve its individual predictive capability. The authors feel, though, that it is not necessary to modify a combined metric (like Joint Nordic Method), since using individual metrics for loudness and for tonality in the proposed regression model herein demonstrated high correlations to the annoyance responses on their own.

4 SUMMARY AND CONCLUSION

The purpose of this study was to investigate how noise signals with varying degrees of prominent tones, similar to those produced by building mechanical equipment, affect subjective annoyance perception and task performance and to develop a prediction model of annoyance using current noise metrics. Subjects completed Sudoku puzzles and a questionnaire modified from the NASA task load index to quantify the overall workload caused by building mechanical noise in this study. No statistically significant effect was found between the tonal signals used in this study and task performance, although there was a trend of decreasing accuracy with increasing tone strengths, based on correct Sudoku puzzle answers. The validity of the modified task load index questionnaire was high based on its reliability coefficient and test–retest coefficient and the average response from the questionnaire was found to significantly correlate with perceived annoyance and loudness of the background noise signals. A factorial repeated measure ANOVA revealed that participants felt more annoyed with increasing background noise level, lower tone frequency and higher tone strength. Correlation analysis with noise metrics and subjective perception ratings found that ANSI loudness level among all other loudness metrics correlates most strongly with annoyance perception, while assorted tonality metrics showed relatively weaker but still statistically significant correlations with annoyance. A statistically significant multivariate regression model with ANSI loudness level and tonal audibility has been developed, which demonstrates an R^2 value of 0.962.

While noise sensitivity of test subjects was surveyed, no statistically significant relations between perception or performance results and noise sensitivity were found, likely due to the limited number of test subjects. Future work in this area is suggested with more test subjects and more tonal signals, to understand better the role of noise sensitivity. Also, tonal noises from actual building mechanical systems often demonstrate multiple tones which may be inharmonic or which can fluctuate in time; additional investigations using tonal signals that incorporate these other factors are recommended.

5 ACKNOWLEDGMENTS

Many thanks to the undergraduate research assistants, Adam Steinbach and Kristin Hanna, who assisted with the study design, subjective testing and data analysis. Partial funding for this project came from an Institute of Noise Control Engineering Undergraduate Research Grant.

6 REFERENCES

1. International Organization for Standardization (ISO), "Acoustics—Description, Measurement and Assessment of Environmental Noise—Part 2: Determination of Environmental Noise Levels", International Standard ISO 1996–2:2007, International Organization for Standardization, (2007).
2. "Seattle Municipal Code Chapter 25.08 Noise Control", Seattle, (2007).
3. "Noise Control Ordinance of the County of Los Angeles", Los Angeles County, (1978).
4. "Ordinance of Minnesota Chapter 389. – Noise", Minnesota, (2008).
5. N.Y. ADC. LAW § 27-770 : NY code – Section 27-770, "Noise Control of Mechanical Equipment", New York, (2006).
6. R.P. Hellman, "Growth rate of loudness, annoyance, and noisiness as a function of tone location within the noise spectrum", *J. Acoust. Soc. Am.*, **75**(1), 209–218, (1984).
7. R.P. Hellman, "Perceived magnitude of two-tone-noise complexes: loudness, annoyance, and noisiness", *J. Acoust. Soc. Am.*, **77**(4), 1497–1504, (1985).
8. S. More and P. Davies, "Human responses to the tonalness of aircraft noise", *Noise Control Engr. J.*, **58**(4), 420–440, (2010).
9. K.H. Lee, P. Davies and A.M. Surprenant, "Quantification of the tonal prominence of complex tones in machinery noise", *NoiseCon04*, (2004).

10. A. Hastings, H. Lee, P. Davies and A.M. Surprenant, "Measurement of the attributes of complex tonal components commonly found in product sound", *Noise Control Engr. J.*, **51**(4), 195–209, (2003).
11. E.E. Ryherd and L.M. Wang, "Implications of human performance and perception under tonal noise conditions on indoor noise criteria", *J. Acoust. Soc. Am.*, **124**(1), 218–226, (2008).
12. P. Susini, S. McAdams, S. Winsberg, I. Perry, S. Vieillard and X. Rodet, "Characterizing the sound quality of air-conditioning noise", *Appl. Acoust.*, **65**(8), 763–790, (2004).
13. B. Berglund, P. Hassmén and A. Preis, "Annoyance and spectral contrast are cues for similarity and preference of sounds", *J. Sound Vibr.*, **250**(1), 53–64, (2002).
14. U. Landström, E. Åkerlund, A. Kjellberg and M. Tesarz, "Exposure levels, tonal components, and noise annoyance in working environments", *Environ. Int.*, **21**(3), 265–275, (1995).
15. U. Landström, A. Kjellberg and M. Byström, "Acceptable levels of sounds with different spectral characteristics during the performance of a simple and a complex non-auditory task", *J. Sound Vibr.*, **160**(3), 533–542, (1993).
16. D. Laird, "The influence of noise on production and fatigue, as related to pitch, sensation level, and steadiness of the noise", *J. Appl. Psychol.*, **17**(3), 320–330, (1933).
17. J.V. Grjmaldi, "Sensori-motor performance under varying noise conditions", *Ergonomics*, **2**(1), 34–43, (1958).
18. K. Holmberg, U. Landstrom and A. Kjellberg, "Effects of ventilation noise due to frequency characteristic and sound level", *J. Low Freq. Noise Vib.*, **12**(4), 115–122, (1993).
19. International Organization for Standardization (ISO), "Acoustics—Assessment of noise annoyance by means of social and socio-acoustic surveys", International Standard ISO/TS 15666: 2003, International Organization for Standardization, (2003).
20. T.H. Pedersen, "The "Genlyd" noise annoyance model: dose-response relationships modelled by logistic functions", DELTA, Hørsholm, Danmark, (2007).
21. C. Marquis-Favre, E. Premat and D. Aubrée, "Noise and its effects—a review on qualitative aspects of sound. Part II: noise and annoyance", *Acta Acust. united with Acust.*, **91**(4), 626–642, (2005).
22. American National Standards Institute (ANSI), "Acoustics—Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment—Part 1 : Determination of Sound Power Level and Emission Sound Pressure Level", American National Standards Institute ANSI/ASA S12.10-2010/Part 1, American Society of America, (2010).
23. W. Aures, "The sensory euphony as a function of auditory sensations", *Acustica*, **58**(5), 282–290, (1985).
24. American National Standards Institute (ANSI), "Procedure for the Computation of Loudness of Steady Sounds", American National Standards Institute ANSI S3.4-2007, Acoustical Society of America, (2007).
25. "Acoustics—Method for Calculating Loudness Level", International Standard ISO 532:1975, International Organization for Standardization, (1975).
26. E. Zwicker, "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)", *J. Acoust. Soc. Am.*, **33**(2), 248–248, (1961).
27. B.R. Glasberg and B.C.J. Moore, "Prediction of absolute thresholds and equal loudness contours using a modified loudness model", *J. Acoust. Soc. Am.*, **120**, 585–588, (2006).
28. Federal Aviation Administration, "Part 36. Noise Standards: Aircraft Type and Airworthiness Certification", Federal Aviation Regulations, Federal Aviation Administration, (1969).
29. Air-Conditioning, Heating, and Refrigeration Institute (AHRI), "Sound Quality Evaluation Procedures for Air-Conditioning and Refrigeration Equipment", Air-Conditioning, Heating, and Refrigeration Institute AHRI/ANSI 1140-2012, Air-Conditioning, Heating, and Refrigeration Institute, (2012).
30. W.E. Blazier, "Revised noise criteria for application in the acoustical design and rating of HVAC systems", *Noise Control. Engr.*, **16**(2), 64–73, (1981).
31. F. Faul, E. Erdfelder, A. Buchner and A.-G. Lang, "Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses", *Behav. Res. Methods*, **41**, 1149–1160, (2009).
32. M. Schutte, A. Marks, E. Wenning and B. Griefahn, "The development of the noise sensitivity questionnaire", *Noise Heal.*, **9**(34), 15, (2007).
33. J. Francis, J. Lee, A. Steinbach and L.M. Wang, "Determining annoyance thresholds of tones in noise", *ASHRAE Trans.*, **120**, (2014).
34. J.W. Grabbe, "Sudoku and working memory performance for older adults", *Act. Adapt. Aging*, **35**(3), 241–254, (2011).
35. H.S. Chang and J.M. Gibson, "The odd–even effect in Sudoku puzzles: effects of working memory, aging, and experience", *Am. J. Psychol.*, **124**(3), 313–324, (2011).
36. "NASA TLX: Task Load Index", [Online]. Available at: http://humansystems.arc.nasa.gov/groups/tlx/, (accessed September 9, 2015).
37. R.F.S. Job, J. Hatfield, N.L. Carter, P. Peploe, R. Taylor and S. Morrell, "General scales of community reaction to noise (dissatisfaction and perceived affectedness) are more reliable than scales of annoyance", *J. Acoust. Soc. Am.*, **110**(2), 939–946, (2001).