PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Sound Quality Evaluation of Refrigerated Truck Noise

Weonchan SUNG[1]; Patricia DAVIES[2]; J. Stuart BOLTON[3]

[1,2,3] Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, U.S.A.

ABSTRACT

Noise from refrigeration units on trucks can be a problem, particularly when the trucks are parked near residences. The development of a sound evaluation method that takes into account the strength of all sound attributes affecting people's responses, including level, is described here. Such a model, coupled with a sound prediction methodology, would be helpful to unit designers who are concerned with optimizing a unit's acoustical performance. The analysis of human subject responses to a variety of refrigerated truck sounds (in the third of a series of three subjective tests) is described. Sixty participants rated 25 recorded and digitally modified refrigerated truck sounds as well as 25 residential HVAC equipment sounds. Fourteen sound quality metrics were evaluated as potential variables in a model to predict the average annoyance ratings of the refrigerated truck sounds. A rate-change-of-level metric that captured the impulsive character of the refrigerated truck unit noise was developed, and the best performing model comprised it, a level metric and a spectral balance metric. The performance of the model when used with signals and ratings taken from other tests is described.

Keywords: Sound Quality, HVAC

1. INTRODUCTION

In many countries, refrigerated trucks move goods that are susceptible to spoilage if they are not under a certain temperature. Refrigerated trucks can generate noise that affects people in nearby residences, whether the vehicle is moving or stationary with the refrigeration unit running (1). In addition, reducing noise from a truck reduces driver fatigue, helps drivers to identify dangerous situations, and can reduce pedestrian and bike-related accidents (2). Here the investigation is on the development of criteria to guide noise control for refrigerated truck units, and to improve their sound quality.

The main sound generating components of refrigerated truck units are fans, compressors, and a diesel engine. In Figure 1 is shown a typical power spectral density generated from a measurement close to a truck refrigeration unit. It can be seen that the sounds consist of harmonic families from the engine (f_0= 45 Hz), compressor (f_0= 30 Hz), as well as a broadband component. The harmonic components strongly affect the sound characteristics and therefore the sound quality of the refrigerated truck unit. Sounds with high amplitude tones or high levels of broadband components may both be loud, but the former increases the tonal character of the sound and the latter decreases it. The distribution of sound energy across the spectrum, i.e., spectral balance, affects the quality of the sound. Two closely-spaced tones can produce a beating phenomenon where the modulation frequency is controlled by the separation of the two tones. When this separation is around 1-8 Hz, variations in level are noticeable, and when it is around 40-80 Hz the level fluctuations are not noticeable but they give the sound a rough character. If a short transient event is repeated continuously at regular intervals, it may be perceived as impulsive or pounding. In many studies researchers have noted that various sound characteristics can make sounds more annoying or more pleasant than would be predicted by considering level alone (3-8).

Although many sound quality studies have been conducted and various models or metrics proposed that take into account various sound attribute strengths, A-weighted sound pressure level is still widely

[1] sung26@purdue.edu

used as the only measure to quantify the sound. Sometimes, this is because the software to estimate the sound attributes' strengths is not available, and sometimes it is because the proposed metrics do not work well. The latter problem may be due to sounds from only a small number of types of units being used for the sound quality model development, thus limiting the applicability of the developed models. Sound attributes that may be important in general for a set of machine types, may not have been varied sufficiently in the sounds used in the model development, or may have co-varied with other sound attributes. In these cases, a proposed sound quality model may not apply to units with a wider range of sound characteristics where attributes no longer co-vary and there is a greater variety of levels of important sound attributes. A few models for diesel engine sound quality have been developed and are used (7-11), however, they are usually tailored to only one application and cannot be applied to another.

Reported in this paper is a study that builds on some previous research (12-14). The experiment described is one of three experiments conducted to develop annoyance models of HVAC&R equipment sounds. Various experimental methods and sounds were used to study people's reaction to a large set of HVAC&R equipment sounds. The results presented in this paper are limited to refrigerated truck unit sounds.

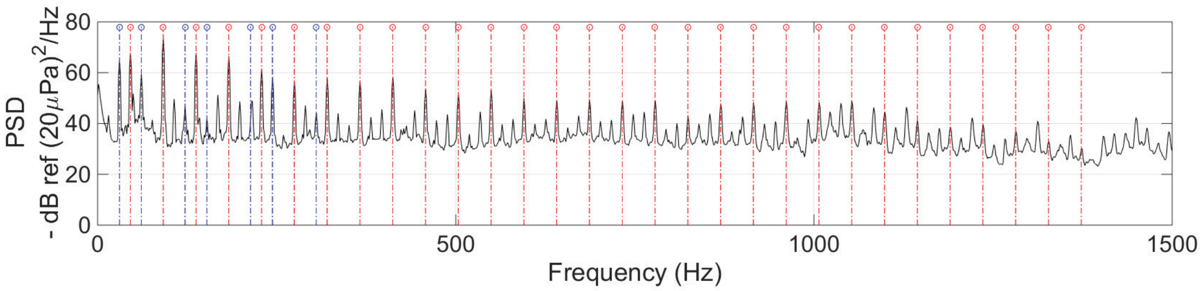

Figure 1 – Example power spectral density for a refrigerated truck unit. The red and blue dotted lines with circles indicate the engine and compressor harmonics, respectively.

2. THE TEST DESCRIPTIONS

In previous tests (12-14), sounds with a wide range of loudness levels (10-50 sones) were presented to subjects. This may have affected the way that subjects weighted the importance of sound attributes other than loudness. In practice, other sound characteristics may be more important because the equipment under examination will compete with equipment types of similar size, and so the loudness variations across units and operating conditions will not be so large within in this set of competing equipment. In this test there were three parts: Parts A, B and C. Part A included quieter sounds (15-35 sones), Part B included louder sounds (25-45 sones), and Part C consisted of a wider range of loudness sounds (10-50 sones) as was used in previous tests. Half of the subjects, Group 1, evaluated Part A sounds first then Part B and Part C, and the other half, Group 2, evaluated Part B sounds first then Part A and Part C. Purdue Institutional Review Board approval was received before testing began (IRB number 1507016324).

2.1 Procedures

When subjects arrived at the designated area, they were given an oral overview of the test. Then, they read and signed the Consent Form, if they wanted to continue. Before starting the actual test, subjects' hearing threshold levels were checked. If their hearing thresholds in the octave bands between 125 Hz to 8000 Hz were 20 dB or below, they could continue with the test. If their thresholds were higher than 20 dB, they were paid $5 and did not continue to take the test.

Subjects read the following scenario before the test: 'While you are listening, it may be helpful to imagine yourself in your garden, at any time during the day or evening, hearing these sounds continuously." Before starting each part of the test, subjects listened to 10 sounds which were chosen to cover the range of the sound characteristics in that part of the test. Also, in each part, subjects practiced rating two sounds, then rated the 50 sounds presented. There were 3-minute breaks between each part. The order of sounds was randomly chosen for each part and for each subject. After all parts were completed, the subject's hearing thresholds were measured once again, and the subjects were asked for comments on the experiment. Finally, the subjects were paid $10.

2.2 Test Environment and Playback

The test was conducted one subject at a time. The test sounds were played using a LynxONE sound card, a Furman SP-20AB stereo amplifier, and Etymotic ER-2 research earphones. The subject had access to a screen, keyboard and mouse, but the computer was outside the double walled IAC sound booth where the subject took the test. The researcher used software to run each part of the test, which controlled the playback of the sounds and automatically recorded the responses. The subjects evaluated the sounds by making a mark on a scale by using the mouse, after the sound was played. The scale had five marks labeled "not at all annoying", "slightly annoying", "moderately annoying" and "very annoying". In the analysis, these marks were assigned values of 2.0, 3.5, 5, 6.5 and 8.0, respectively. Values from 1 to 9, with resolution controlled by the resolution of the screen, were possible outcomes because subjects could mark anywhere on the line which extended beyond the first and last marks.

2.3 Test Sounds

The tests included the sounds of both refrigerated truck units and residential units. In the test, both recordings and modified recordings were used. The sounds were modified to include many different sound attribute levels and to reduce correlation between metrics across the set of sounds presented. Thus, the correlations between sound quality metrics related to level, tonalness, spectral balance, and fluctuation were significantly lower than when only the original recordings were included. A total of 120 sounds were used in the test, with each part containing 50 sounds, with 15 of the sounds included in all three parts of the test. These *common sounds* were included to allow adjustment of the annoyance ratings if the subjects used the rating scales slightly differently from part to part. Table 1 contains information on the sounds used in the tests. Part A was the quieter test, mostly consisting of residential sounds, and Part B was the louder test, mostly consisting of refrigerated truck sounds.

Table 1 – Summary of the test signals

Part	Signals
Part A (Quieter, mostly residential sounds)	50 Sounds 28 original, 22 modified 36 residential, 14 refrigerated truck
Part B (Louder, mostly refrigerated truck sounds)	50 Sounds 30 original, 20 modified 11 residential, 39 refrigerated truck
Pact C (Wider Loudness range)	50 Sounds 19 original, 31 modified 24 residential, 26 refrigerated truck

2.4 Subjects

The subjects were recruited through Purdue online bulletin boards and flyers posted on the Purdue campus and in local community buildings. Of the 62 subjects who volunteered to take the test, 60 passed the hearing test. Thirty of them identified themselves as male and 30 as female. The subjects ages were from 18 to 62 years, the average age was 28.4 years and the median was 26 years. Thirty-two test participants grew up in the U.S., 25 subjects grew up in Asia (15 East Asia, 10 India), 1 subject grew up in South America, and 2 subjects grew up in Africa. Most of the subjects has no knowledge of sound quality and noise control. Five of the participants have previously participated in similar experiments, and 3 sometimes participate in musical events.

3. RESULTS AND DISCUSSION

The average of the subjects' annoyance ratings and the standard deviation of estimated mean (standard error) were calculated for each of the two subject groups and for each part of the experiment. Test results, correlation between average annoyance ratings and sound metrics, sound metric adjustments, and performance of various linear annoyance models are described in this section.

3.1 Average Annoyance Ratings

The average annoyance ratings for the signals are shown in Figure 2. Results for Group 1 and Group 2 subjects are shown separately. Group 1 subjects (red) heard the quieter sounds (Part A) first, while the Group 2 subjects (blue) heard the louder sounds (Part B) first. Part A sound ratings from Group 2 subjects were slightly lower than those from Group 1 subjects, and Part B sound ratings from Group 1 subjects were slightly higher than those from Group 2 subjects. This indicates that the subjects' responses were affected by the order of the parts in the experiment. Trends in signal ratings are similar. Part A and Part B ratings were adjusted so the ratings of the common sounds in Parts A, B and C matched (on average). The adjustments were made by first fitting a straight line to the Part C versus Part A (or Part B) responses for the common sounds, and then using the estimated intercepts and gradients of the lines to adjust the averaging ratings of all the signals in Part A and Part B. These adjusted Part A and Part B ratings were used along with Part C ratings in the model development.

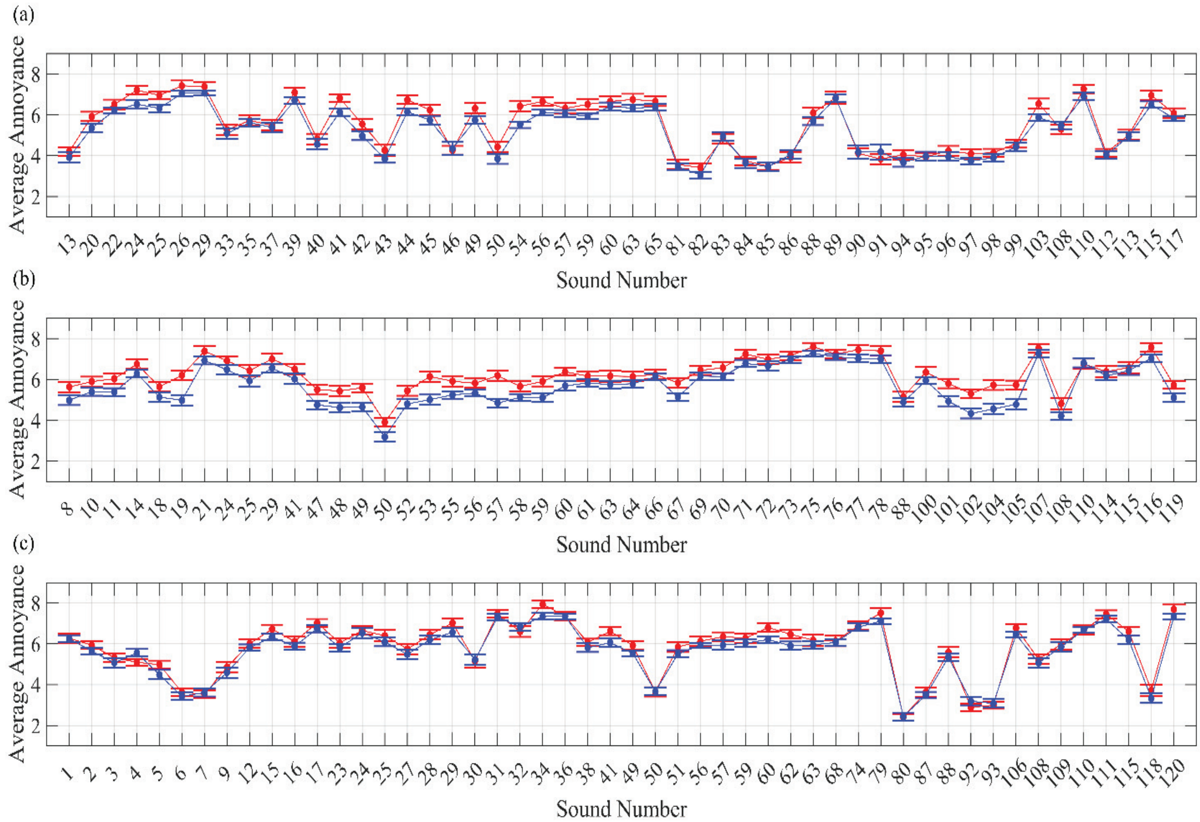

Figure 2 – Average annoyance rating with standard error bars. (a) Part A ratings – quieter test, (b) Part B ratings – louder test, and (c) Part C ratings – wider loudness range test. Red is for Group 1: Part A→B→C, blue is for group 2: Part B→A→C.

3.2 Metrics and Adjustment of Metrics

The sound metrics examined are listed in Table 2. Most of the metrics were calculated by using Head Acoustics' ArtemiS software and some were programed and calculated using MATLAB. The thresholds for sound metrics were calculated assuming that while the different values of the metric at the lower end of the metric range do reflect differences in the sound characteristics, these differences did not greatly affect the annoyance ratings for those sounds. Adjusted Tonality and adjusted Sharpness were calculated by thresholding Tonality and Aures' Sharpness. When calculating the thresholds for Tonality and Aures's Sharpness, the average annoyance results of both residential units and refrigerated truck units were used even though the annoyance prediction models were developed separately. The R^2 values for four-metric linear models (metrics: N_5, S_{A5adj}, T_{5adj}, R_5) with different adjustment thresholds were calculated and the thresholds that resulted in the highest R^2 values were selected. The Sharpness and Tonality thresholds for this model were thus set at 2.5 acum and 0.27 tu, respectively. If the value of Tonality or Aures' Sharpness is less than the threshold, the adjusted metric is zero, and if it is larger, it is set to the *original metric value – the threshold*.

In the second test (in the series of tests leading to the experiment described here), which was a semantic differential test, the fourth attribute in the factor analysis of the responses to refrigerated truck sounds was aligned with the ratings on the "Impulsiveness" scale (14). Also, in the first of this series of tests in which subjects wrote down words or sentences to describe the sounds, many subjects used words and descriptions that are related to impulsiveness, when describing sounds with diesel engine components (13). So, in addition to the commonly considered sound metrics for annoyance models, several impulsiveness-related metrics were also considered. The most effective impulsiveness metric (when included in the annoyance models) was the maximum rate of change of the loudness (*RCL*), a metric that has been used in assessment of transient environmental sounds (15). To calculate *RCL*, Zwicker loudness time histories were calculated by using the Head ArtemiS software and then a 15-point digital FIR filter (designed using the *firpm* program in MATLAB) was used to calculate the derivative of the loudness time history. The maximum rate of change of loudness was used to quantify the impulsiveness characteristic.

Table 2 – Sound metrics examined in the test. The "adj" indicates that the metric value is adjusted.

Metric Symbol	Metric Description	Threshold for Adjusted Metric
N_5	Zwicker time varying Loudness exceeded 5% of the time	N/A
S_{VB5}	Von Bismarck Sharpness exceeded 5% of the time	N/A
$S_{A5} \rightarrow S_{A5adj}$	Aures Sharpness exceeded 5% of the time	2.5 acum
$T_5 \rightarrow T_{5adj}$	Tonality metric exceeded 5% of the time	0.27 tu
K, RCL	Impulsiveness – Kurtosis and Loudness derivative	N/A
R_5	Roughness (16) exceeded 5% of the time	N/A
FS_5	Fluctuation Strength (16) exceeded 5% of the time	N/A
dBA, dBC	A or C weighted sound pressure level	N/A
SQI*	SQI (6) calculated from sound pressure not sound power	N/A
PA	Psychoacoustic Annoyance (16)	N/A

3.3 Annoyance Models

The annoyance prediction models for the refrigerated truck sounds, are a linear combination of some of the sound metrics shown in Table 2. All combinations of metrics were considered and the models with the highest R^2 values were selected, but models where each metric represented strengths of different sound attributes were preferred, because those attributes could be directly related to signal characteristics, and thus to components in the machine. The average of the annoyance ratings plotted against annoyance predictions for the best 1, 2 and 3 metric models are shown in Figure 3. Here responses from all three parts of the test were used to estimate the model coefficients. The most accurate single-metric predictor is N_5, see Figure 3(a). Inclusion of the adjusted Aures' Sharpness metric in the model increases the R^2 value by 0.02 (Figure 3(b)). Unlike the residential annoyance prediction model, see (17) for details, including the adjusted Tonality and Roughness metrics in the model did not improve annoyance predictions for the refrigerated truck sounds. However, by including the *RCL* metric as the third metric in the model, the R^2 value increased by 0.01 (Figure 3(c)). The three models estimated using only the 26 responses to refrigerated truck sounds in Part C of the test were used to predict the responses to refrigerated trucks sounds in Part A (14 sounds) and Part B (39 sounds) of the test. The results are shown in Figure 4. Responses to Part A and Part B sounds are accurately predicted by the Part C model. Inclusion of S_{A5adj} and *RCL* in the model helped to improve the prediction accuracy. From Figure 4 it can be seen that when the three-metric model was used, Part B sounds were noticeably closer to the one-to-one line (black) than when using the two-metric model.

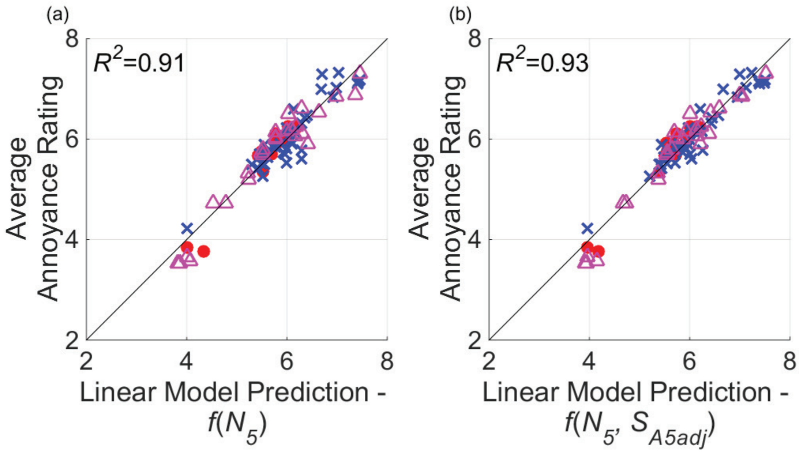

Fig. 3 – Average of annoyance ratings of refrigerated truck unit sounds plotted against annoyance predictions. The standard deviations of the estimated average annoyance range between 0.12 And 0.23. The markers indicate Part A (●), Part B (x), and Part C (△) sounds.

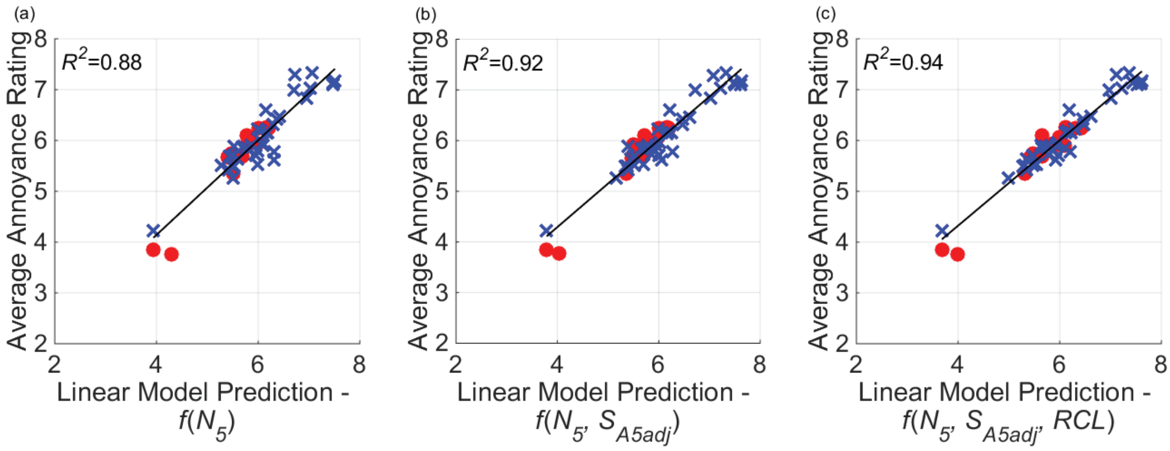

Fig. 4 – Average of annoyance ratings in Part A and Part B plotted against annoyance predictions from models developed using responses in Part C only. The markers indicate Part A (●) and Part B (x) sounds

4. CONCLUSIONS AND DISCUSSIONS

The third in a series of tests focused on evaluation of HVAC&R equipment noise has been described here. Subjects rated sounds in the three parts of this test. Part A contained ratings of quieter sounds, Part B louder sounds, and Part C sounds that spanned a wide loudness range. There were only slight differences in the results from the group that did Part A first and the Group that did Part B first. In both cases there tended to be slight overshoot. The group that heard the quieter sounds first rated the louder sounds higher than the group that heard the louder sounds first. Similarly the group that heard the louder sounds first rated the quieter sounds lower than the group that heard the quieter sounds first. The performance of one-, two- and three-metric models when predicting the annoyance ratings of refrigerated truck unit sounds was examined. It was found that results could be improved by adjusting two of the metrics so that they only played a role in the model above a certain threshold value. The most important metric for an annoyance prediction was Zwicker Loudness exceeded 5% of the time (N_5). The second most important metric was adjusted Aures' Sharpness, and inclusion of the maximum rate of change of the loudness (*RCL*) in the model was also significant, particularly for predicting the annoyance of the louder sounds. The models estimated using responses to Part C sounds only, also predicted the responses to sounds in Part A and Part B very accurately.

ACKNOWLEDGEMENTS

Thanks to D.J. Carr at the Ray W. Herrick Laboratories for help in the design and preparation of this test. Also thank to industry sponsors for their support, insights and advice.

REFERENCES

1. Thiessen GJ, Lapointe AC. Effect of intermittent truck noise on percentage of deep sleep. *Journal of the Acoustical Society of America*. 1978;64(4):1078-80.
2. Wang X, Zhou Y, Geovaers R, Holguin-Veras J, Wojtowicz J, Campbell S, Miguel J, Webber R. Feasibility of installing noise reduction technologies on commercial vehicles to support off-hour deliveries. New York State Energy Research and Development Authority, New York, U.S.A., 2014
3. Töpken S, Van de Par S. Perceptual Dimensions of Fan Noise and Their Relationship to Indexes Based on the Specific Loudness. *Acta Acustica united with Acustica*. 2019; 105(1):195-209.
4. Yamaguchi T, Minorikawa G, Kihara M. Study on evaluation method of the pure tone for small fan. *Proceedings of INTER-NOISE 2014.*16-19 October 2014; Melbourne, Australia, 2014;249(6):1629-1633.
5. Wang SY. Determination of sound quality of refrigerant compressors. *Proceedings of the 1994 International Compressor Engineering Conference*. 19-22 July 1994; Purdue University, West Lafayette, Indiana, U.S.A.; 241-246.
6. AHRI/ANSI, AHRI/ANSI Standard 1140: Sound Quality Evaluation Procedures for Air-Conditioning and Refrigeration Equipment. Air-Conditioning, Heating, and Refrigeration Institute/ American National Standards Institute, Virginia, U.S.A., 2012.
7. Hasting AL, *Ph.D. Thesis: Sound Quality of Diesel Engines*, Purdue University, West Lafayette, Indiana, U.S.A., 2004.
8. Bodden M, Heinrichs R. Diesel sound quality analysis and evaluation. *Proceedings of Forum Acusticum 2005*. August 29 –September 2 2005; Budapest, Hungary;1-4.
9. Ingham R, Otto N, McCollum T. Sound quality metric for diesel engines. *SAE transactions*. 1 Jan 1999; 108(6); 3003-3007.
10. Sellerbeck P, Nettelbeck C, Heinrichs R, Abels T. Improving diesel sound quality on engine level and vehicle level-a holistic approach. *SAE Technical Paper*. 15 May 2007.
11. Liu H, Zhang J, Guo P, Bi F, Yu H, and Ni G. Sound quality prediction for engine-radiated noise. *Mechanical System and Signal Processing*. May 2015; 56; 277-287
12. Sung W, Davies P, Bolton J.S. A methodology to modify steady state heating, ventilating, air conditioning and refrigerant equipment noise. *Proceedings of NOISE-CON 2016*. 13-15 June 2016; Providence, Rhode Island, U.S.A.; 299-306.
13. Sung W, Davies P, Bolton J.S. Descriptors of sound from HVAC&R equipment. *Proceedings of NOISE-CON 2017*. 12-14 June 2017; Grand Rapids, Michigan, U.S.A.; 580-588.
14. Sung W, Davies P, Bolton J.S. Results of a semantic differential test to evaluate HVAC&R equipment noise. *Proceedings of INTER-NOISE 2017*. 26-29 August 2017; Hong Kong, China; 2390-2398.
15. Marshall AJ, *Ph.D. Thesis: Development of a Model of Startle Resulting from Exposure to Sonic Booms*, Purdue University, West Lafayette, Indiana, U.S.A., 2012.
16. Zwicker E, Fastl, *Psychoacoustics: Facts and Models*. Springer Science & Business Media; 4th Edition, 2013 Mar 14.
17. Sung W, Davies P, Bolton J.S. Sound quality evaluation of residential HVAC&R equipment. *Proceedings of INTER-NOISE 2018*. 26-29 August 2018; Chicago, Illinois, U.S.A.; 1066-1074.