NSA-2015 Goa
National Symposium on Acoustics
"Acoustics for Ocean Environment"

STATISTICAL PROCESSING OF SUBJECTIVE TEST DATA FOR SOUND QUALITY EVALUATION OF AUTOMOTIVE HORN

Tapan K. Mahanta, B. Venkatesham and B. Karri

Department of Mechanical & Aerospace Engineering, Indian Institute of Technology Hyderabad, Telangana, India-502025.

Corresponding author's email: venkatesham@iith.ac.in

Product sound quality has a significant role in buying decision and customer satisfaction. An often used method to assess the sound quality of any product or equipment is a subjective listening test where the sound is heard by a panel of subjects (jury) who then rate the sound quality. Subjects use a semantic differential rating wherein they evaluate the presented sounds based on a bipolar variable. The two extremes of the rating scale are labeled with an adjective and its antonym respectively. In the present study, a subjective listening test has been conducted to assess sound quality of automotive horns. The data obtained are then analyzed using statistics to gain insights. Twenty two horn sound samples were judged by thirty participants aged 20-40 years who had normal hearing. Binaural head set (BHS) instrument was used to record horn sound samples in open ground (neglecting wind noise effect). Sounds are recorded two meter from horn in front direction and used for subjective test. For the subjective test and subsequent statistical analysis, a four step procedure has been used. In the first step, the participants were asked to rate the sound quality for each horn based on seven bipolar variables. These bipolar variables are soft/loud, calm/frightening, slow/fast, relax/tense, safe/danger, vague/distinct and pleasant/unpleasant. For each bipolar variable, a seven verbal interval scale was used ranging from one extreme to another in degree, for example extremely pleasant to extremely unpleasant. In the present study, only one bipolar variable namely pleasant/unpleasant has been chosen for further statistical analysis. The same analysis can, however, be extended to remaining variables. In the second step to account for differences in individual rating, the magnitude data obtained was normalized based on the median rating for a particular horn. In the third step, two way ANOVA analysis has been applied to evaluate the statistical significance of differences due to subjects and/or horns. The normalized data indicated that the differences due to subjects were not statistically significant. In the fourth step, the twenty two horn sounds are classified based on Tukey's method into groups wherein horns in each group have similar sound quality. Tukey's analysis involves estimation of studentised quartile range based on percentage of confidence interval, error degree of freedom and number of groups' degree of freedom.

1. Introduction

An automotive horn is used by the driver to warn a fellow road-user about a potential danger and avoid accidents. However, the noise from such horns particularly in big cities has a negative influence on the living conditions of people [1]. For example: three major cities (Delhi, Mumbai, and Kolkata) in India are listed in the top ten noisiest cities of the world and a major source of noise is from honking. So, manufacturers are interested to improve the sound quality of horns without compromising basic functional requirements. Sound quality can be assessed through objective and subjective studies. Objective analysis calculates psycho-acoustic parameters and indices like sound pressure level, loudness, roughness, sharpness and fluctuation of strength, pleasantness index etc from recorded sound samples. Subjective analysis requires a subjective listening test where the sound is heard by a panel of subjects (jury) who then rate the sound quality. Subjects use a semantic differential rating wherein they evaluate the presented sounds based on a bipolar variable. The two extremes of the scale are labeled with an adjective and its antonym respectively. Afterwards, different statistical analyses can be conducted to gain insights from the subjective test data. This paper describes a statistical procedure for classification of horn data based on pleasantness characteristics.

2. Methodology for subjective listening test and statistical processing

The methodology used for the listening test and subsequent statistical processing is shown as a flow chart in Figure 1.

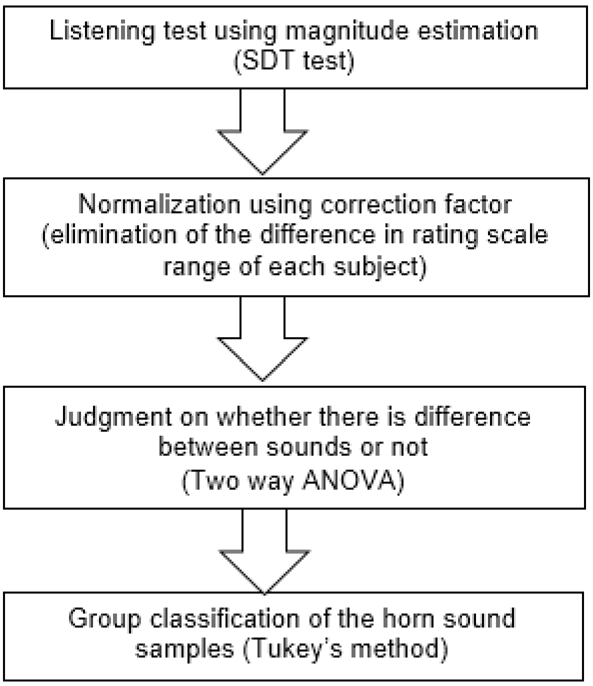

Figure 1. Flow chart used for statistical processing of listening test

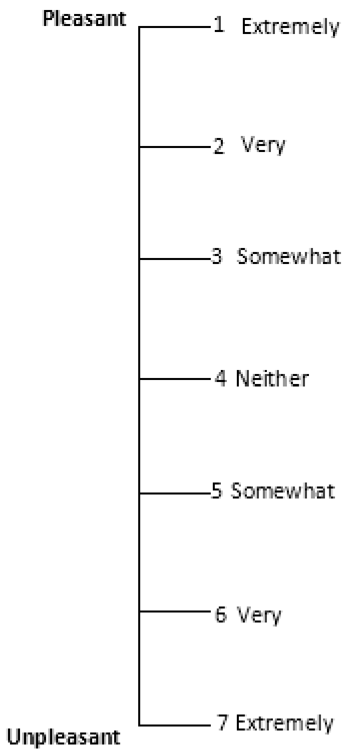

Figure 2. Verbal Interval scale in SDT Test

The various steps in the above chart are discussed below.

2.1 Listening test using magnitude estimation (SDT test)

A total of twenty two horn sound samples were recorded with binaural head set (BHS) as described in Venkatesham et al [2]. The recording was done by using equipment shown in Figure 3. Among two channels from BHS data for each horn sample, only the channel which has the largest overall sound pressure level was chosen for jury test. A jury comprising of thirty members in the age group of 20-40 with normal hearing participated in semantic differential evaluation method following the guidelines given by Otto et al.[3]

Semantic differential test (SDT) consists of seven bipolar variables which are soft/loud, calm/frightening, slow/fast, relax/tense, safe/danger, vague/distinct and pleasant/unpleasant. For each bipolar variable, a seven verbal interval scale was used ranging from one extreme to another in degree, for example extremely pleasant to extremely unpleasant as shown in Figure 2.

An interactive GUI was designed to conduct the jury test. It allows the subjects to listen to each sample as many times as required to make a judgment. They then rate the level of pleasantness for each sound on verbal interval scale shown in Figure 4, by moving sliding bar from one (extremely pleasant) to seven (extremely unpleasant). The interface snapshot isshown in Figure 4.

In the present analysis, only one bipolar variable data namely pleasant/unpleasant has been chosen for further statistical processing as described in the subsequent sections.

Figure 3. Jury evaluation experiment test setup

Figure 4. Snapshot of GUI for subjective pleasantness test

2.2 Normalization using correction factor

Environmental and emotional condition play an important role in subjective evaluation of automotive horn sound quality. Data normalization was carried out on the listening test data to minimize cognitive effects. In order to minimize this difference, a correction factor must be applied to each individual rating data which normalizes the data. The correction factor for i_{th} subject (C_i) is defined as [5].

$$C_i = \frac{1}{J}\sum_{j=1}^{J}(x_{ij} - m_j),$$

Where, x_{ij} is the the rating value of the i_{th} subject to the j_{th} sound sample and m_j the median value of the j_{th} sound sample. The next step is to subtract the correction factor from the rating values to the sample sounds for each subject. After completing this step, one can obtain the normalized data, in which the individual difference in rating is eliminated. Normalization of data by using correction factor is shown in Table 1.

Table 1. Normalization of raw data from listening test

Sound No. / Subjects	1	2 …………	J	
Subject 1	$x_{11}-c_1$	$x_{12}-c_1$ ……	$x_{1J}-c_1$	$c_1=\frac{1}{J}\sum_{j=1}^{J}(x_{ij}-m_j)$
Subject 2	$x_{21}-c_2$	$x_{22}-c_2$ …..	$x_{2J}-c_2$	$c_2=\frac{1}{J}\sum_{j=1}^{J}(x_{ij}-m_j)$
……. ………	………	………	…………………	…………………
Subject I	$x_{I1}-c_i$	$x_{I2}-c_i$ ……	$x_{2J}-c_i$	$c_i=\frac{1}{J}\sum_{j=1}^{J}(x_{ij}-m_j)$
Median	m_1	m_2 …………….	m_J	

2.3 Two way ANOVA

The two-way ANOVA compares the mean differences between groups [6]. The current data considered two groups namely subject response and sound sample. Horn sound sample was taken as fixed factor and subject data as random factor. The null hypothesis for a fixed factor is “Sound quality of all horn sound samples are same” and the null hypothesis for a random factor is “There is no influence of each subject on horn sound quality rating”. Table 2 shows the mathematical expressions of the null hypotheses, where, α_i and β_j are mean values for each level of fixed and random factor, respectively.

Table 2: Hypothesis used for the two way ANOVA

	Null hypothesis	Alternative hypothesis
Fixed factor	$H_{0A}=\alpha_1=\alpha_2=\alpha_3=\ldots\ldots=\alpha_I=\mu_r$	H_{aA}: at least one on $\alpha_i \neq \mu_r$
Random Factor	$H_{0B}=\beta_1=\beta_2=\beta_3=\ldots\ldots=\beta_J=\mu_f$	H_{aA}: at least one on $\beta_j \neq \mu_r$

2.4 Classification of Sounds by Tukey Method

Based on Two-way ANOVA analysis, if null hypothesis is rejected then further analysis is needed to know where inequalities exists among the different means. Many methods exists to detect differences between individual means. One of them is Tukey test [6] and it is a multiple comparison test. That is, it can compare many mean values for differences and lead to group classification. The procedure for calculating the studentized quartile range for group classification is as follows.

1. Compute the standard error ($S_{\bar{x}}$)

$$S_{\bar{x}} = \sqrt{\frac{MSE}{n}}$$

Where, MSE is mean square error and n is number of samples.

2. Rank the sample means ($\overline{X_I}$), from lowest to highest.

3. Find range, $|\overline{X_a} - \overline{X_b}|$, For all pair of sample mean
   Where, $\overline{X_a}$ and $\overline{X_b}$ are mean of a and b sample respectively

4. For any two samples a and b. If the value $\frac{|\overline{X_a}-\overline{X_b}|}{S_{\bar{x}}} \leq Q_{\alpha,\nu,k}$ then the two sounds are in the same group. If not then the two sounds are in different group. $Q_{\alpha,\nu,k}$ is the critical values of the q distribution, for the Tukey Test that can be found from appendix table in reference [6]. Where α is the significance level, ν error degree of freedom and k is number of groups.

3. Result and Discussion

Figure 5 shows a comparison of raw data and normalized data calculated according to Section 2.2. It is observed that for normalized data, the range over which the subjective rating value varied was smaller as compared to raw data.After normalization, the two way ANOVA was applied to both raw data and normalized data. It reveals that for raw data (shown in Table 4), both fixed factor (horn sound sample) and random factor (subject) null hypothesis is rejected because in this case $p < \alpha$. In Table 5, for normalized data, it was observed that in case of random factor, the null hypothesis can be accepted because case $p > \alpha$, but on the other hand for fixed factor the null hypothesis should be again rejected due to $p < \alpha$. From this analysis, it is concluded that after normalization, the subject variation in rating is removed while the sound samples have different sound quality pleasantness.

Finally Turkey's method is applied on the normalized data as described in Section 2.4. For this method the calculated standard error is $S_{\bar{x}}$ = 0.218, studentised quartile range $Q_{\alpha,\nu,k}$= 5.081 where α = 0.05, ν = 638 (error degrees of freedom) and k = 22. The valid range w = $Q_{\alpha,\nu,k} * \sqrt{\frac{MSE}{n}} = 1.1$

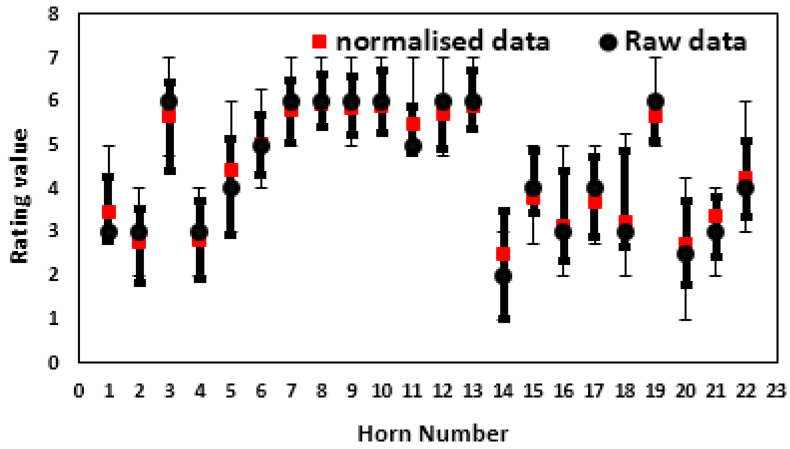

Figure 5. Average values and quartile values related to 25% and 75% for raw and normalized data

Following this, based on steps 2to 4 , the sound samples were formed in to three groups as follows

- Unpleasant group: M3DH6, M5DH2, M3DH4, M3DH5, M3DH3, M2DH1, M5DH1, M2SH1.
- Pleasant group: M1SH1, M2DH2, M1DH1, M2SH2, M4DH1, M1SH2, M2DH5.

The horns M5DH3, M3DH2, M3DH1, M1DH2, M2DH6, M2DH4, and M2DH3 could not be classified in either of pleasant or unpleasant group with confidence.

Where notation followed for different horns used in the test. "M'' stands for manufacturer, DH and SH refer to horn type.

Table 3: Results of two-way ANOVA for raw data

Source of variation	Sum of squares	Degree of freedom	Mean square	F ratio	P value	F value
Subject	588	29	20.27	13.61	6.11E-49	1.49
Automotive horn	966	21	46	30.89	9.87E-82	
Error	907	609	1.49			
Total	2461	659				

Table 4: Results of two-way ANOVA for normalized data

Source of variation	Sum of squares	Degree of freedom	Mean square	F ratio	P value	F value
Subject	4	29	0.14	0.09	1	1.49
Automotive horn	966	21	46	30.89	9.87E-82	1.57
Error	907	609	1.49			
Total	1877	659				

4. Conclusions

A statistical processing method to classify horn sound samples has been discussed. It involves data normalization, Two-way ANOVA, and Tukey's classification method. This methodology has been demonstrated on subjective test data for one bipolar variable namely pleasant/unpleasant. Based on the analysis, horn sound samples which had extreme interval rating could be classified in to pleasant and unpleasant group. However, some sound sample could not be grouped with confidence. It requires further research on classification methods.

References

[1] Waves, Brüel & Kjær, April 2015 Page 16-17.

[2] B.Venkatesham, N. Pradeep, Goutham and Jongdae Kim. Preliminary study on automotive horn sound quality, *National Symposium on Acoustics*, 2012, Tamil Nadu, India, 5-7 December, (2012).

[3] N. Otto, S. Amman, C. Eaton and S. Lake, (2001). Guidelines for jury evaluations of automotive sounds. *Sound and Vibration*, **35**(4), 24-47.

[4] M. Bodden, (1997). Instrumentation for sound quality evaluation. *Acta Acustica united with Acustica*, **83**(5), 775-783.

[5] S. H. Shin, J. G. Ih, and H. Jeong, (2003). Statistical processing of subjective listening test data for PSQ. *Noise Control Engineering Journal*, **51**(4), 232-238.

[6] J. H. Zar, 2010. *Biostatistical analysis*. Pearson Education India. Appendix B, 724-725.

Acknowledgement

We would like to thank Department of Science and Technology, Government of India for financial support to conduct current investigation and Indian Institute of Technology, Hyderabad for providing the required resources.