



Sound Preference Development and Correlation to Service Incidence Rate

Terry Hardesty^{a)}
Sub-Zero, 4717 Hammersley Rd., Madison, WI, 53711, United States

Eric Frank^{b)}
Todd Freeman^{c)}
Gabriella Cerrato^{d)}
Sound Answers, Inc., 6855 Commerce Blvd, Canton, MI 48187, United States

Traditional methods of sound performance measurement for appliances (A-weighted sound power) provide an incomplete picture of the customer's experience. This is often exemplified by the poor correlation between sound power and customer complaints.

In order to satisfy customer demands for better sounding refrigeration appliances, Sub-Zero needed to gain a better understanding of customer's perception of sounds generated by the appliance. This was accomplished by conducting a formal Sound Quality jury evaluation and by correlating objective signal processing parameters (Sound Quality metrics) computed over the sounds presented to the jurors to their subjective ranking. The direct result of the study was a sound preference algorithm utilizing several sound quality metrics (rather than simply sound power) that represented the customer's perception of the sounds.

The data generated by the preference algorithm were then correlated with field data to identify performance acceptance thresholds. This rearward looking analysis provided a basis for establishing market-driven product sound preference targets for new product development.

^{a)} email: terry.hardesty@subzero.com

^{b)} email: eric.frank@soundanswers.net

^{c)} email: todd.freeman@soundanswers.net

^{d)} email: gabriella.cerrato@soundanswers.net

1 INTRODUCTION

The landscape of the luxury refrigeration market has changed significantly over the past decade. Customers' expectations are much higher for aesthetics, performance, reliability, and sound to name a few. Additionally, customers are placing refrigeration in non-kitchen locations such as offices, home theaters, family rooms, and even bedrooms. These new locations elevate awareness of the sounds generated by the appliances.

Traditionally, refrigeration products have been acoustically labelled by using their A-weighted sound power level, and noise and vibration laboratories around the world are equipped with application-specific testing rooms and instrumentation dedicated to the measurement of the sound power spectrum. Original Equipment Manufacturers (OEMs) in the refrigeration industry measure the sound power of their products and give sound power specifications to their suppliers.

As in many applications of Sound Quality, it was determined that this basic method of measuring noise provided an incomplete picture of customer's expectation and satisfaction. The results from informal listening studies did not correlate well with measured sound pressure data. Certain sounds were rated very objectionably even though their Sound Pressure levels were very low. See Figure 1: Sound Preference Score vs. dBA Figure 1.

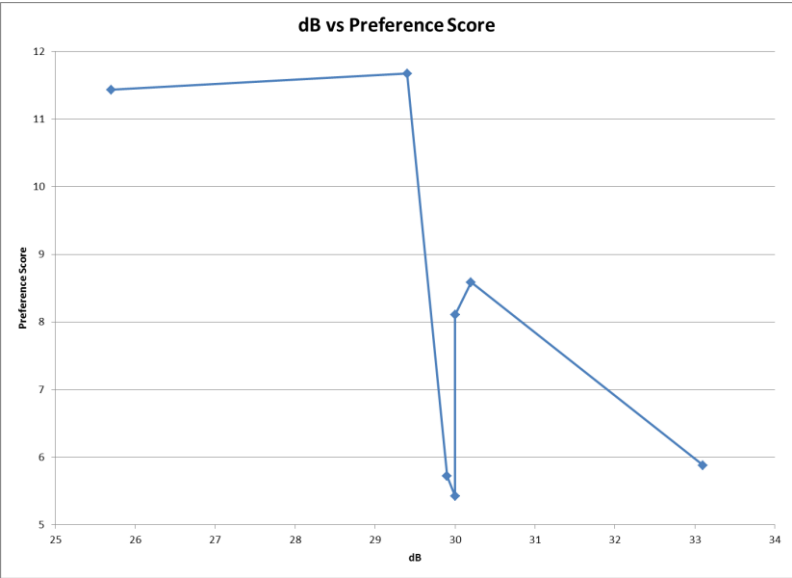


Figure 1: Sound Preference Score vs. dBA

Generally, as a passive appliance, lower overall sound levels are desired for a refrigerator. That said, there are low intensity sounds that can occur in the normal operation of a refrigerator that are particularly objectionable. Based upon these results, a study was conducted to quantify the seemingly subjective human response in order to develop a sound preference algorithm to predict customer satisfaction.

2 CONVENTIONAL METHODS

Historically most refrigeration product sounds have been characterized via sound power tests, overall sound pressure levels with a single microphone, and/or other standardized acoustic tests.

Sound Power is computed from a spatial and temporal average of A-weighted noise measured at different microphone locations. It is great parameter for quantifying the noise radiated from a product in terms of its acoustic power, but this is typically not what we hear. We do not hear “averages”, spatial or temporal, rather we hear features in the sound as they appear instantaneously and these depend on our location.

Customers’ overall impression of a sound, or Sound Quality, is made of their reaction to different features, such as amplitude, tonality, modulation, roughness and so forth. Of all these, only amplitude is captured by Sound Power. In order to understand and quantify Sound Quality it is necessary to understand first which features of the sound most affect the perception and next how to quantify each of these.

3 ACOUSTIC MEASUREMENTS

To assess Sound Quality, one needs to position a binaural head in front of the refrigerator, and record the sound as heard by a customer approaching the refrigerator and opening its doors. This is clearly different than utilization of a single microphone or even utilizing multiple microphones as is done for sound power evaluations.

An artificial binaural head with microphones at the ears’ locations can accurately capture the sound as it would be heard by a human being. The positioning of the microphones at the ear locations provides the spatial cues necessary for proper reproduction, particularly if the environment is reflective, causing the sound waves to be reflected and diffracted prior to reaching the binaural head.

For the purposes of this study, recordings of several representative refrigeration appliances we made using a Head and Torso Simulator positioned to replicate expected consumer conditions. These recordings were taken during normal (cycling) operating conditions (both refrigerator and freezer compartments) of the appliance in a quiet environment replicating the environment utilized by the end user.

4 JURY STUDY

In order to study the customers’ acoustic expectations for this product type, a Sound Quality jury project was conducted where sounds from several different high-end refrigerators were played back to a formal jury for subjective evaluation. The results of the jury were correlated statistically to objective parameters to derive an objective model of refrigerator Sound Quality.

4.1 SQ/Jury Process

The sound quality and jury process can be extremely valuable to study the acoustic opinions of end users, as well as to gain an understanding into the detailed acoustic features that are most significantly affecting these perceptions. To extract the most useful information from a jury group

for this type of study, however, can require significant planning in how the study is conducted. In general terms the process utilized for this study is noted below in Figure 2.

The first steps include conducting acoustic measurements of many product variants. The purpose of this step is to gather a wide range of sounds that capture the entire spectrum of possible sounds that could be presented to the end customer. Often times an informal jury study can then be utilized to determine which of these sounds are most appropriate to present to the formal jury, as well as to begin to form hypothesis for which acoustic features are most likely effecting the subjective preferences of the end users. The final sounds presented to the jury are most often a combination of ‘real’ recorded sounds in addition to sounds that have been digitally modified.

In this process there exist two main tracks, one which is used to obtain the voice of the customer via formal jury evaluations, and a second which involves data analysis to objectively discriminate between each of the sounds presented to the jury. It is through the study of the results of these two tracks that the subjective opinions of the jury can be correlated to specific noise features of the product(s).

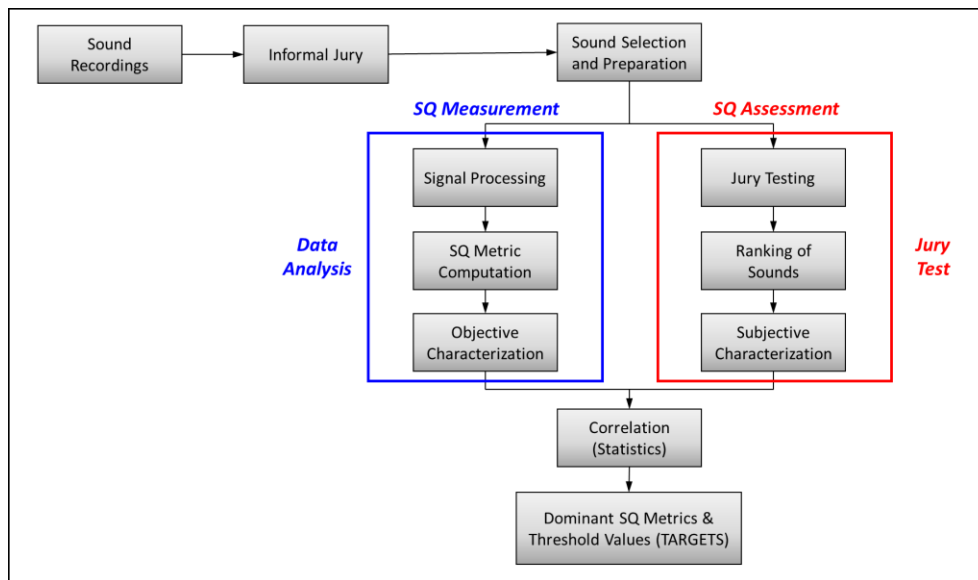


Figure 2: Sound Quality & Jury Process

4.2 Design of Jury Tests

For this study, over 40 recorded sounds were gathered to provide a wide range of product sounds. All measurements were recorded utilizing a head and torso unit in a quiet, representative environment.

The next steps was to characterize each of the refrigerator sounds with appropriate descriptive terms that the jury could understand and relate to. To accomplish this, interviews were conducted in which people were asked to describe sounds (good or bad) that they have heard or expect to hear from a refrigerator. The candidates for this informal study included technical engineering staff along with Regional Distributors who had extensive experience with the operating noise of the refrigerators and direct feedback from customers. Some of the descriptors, like “loud” or “quiet”, were very simple to equate to Sound Quality features, but others, like “whiney”, were less obvious.

Based upon the results from the informal jury study, dozens of refrigerator recordings were analyzed to identify a group of recordings that could encompass the entire expected range of production representative sound attributes. In the end eight recordings were selected to present to the formal jury. Two additional sounds were created digitally to test people's response to very specific sound features in order to fill out the matrix of sounds that would be used for the jury test.

A jury of 53 employees was selected with a concerted effort to have a mixed demographic. Age, gender, and technical expertise were factors considered in selecting jurors. Juror performance was evaluated at the end of the test and a small number were excluded from the data due to unsatisfactory repeatability or consistency results.

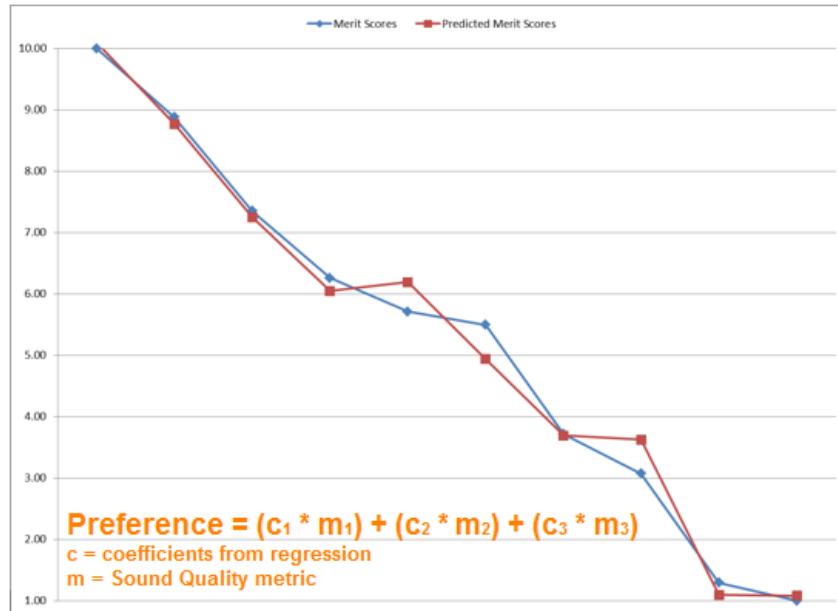
Two types of test are generally administered for Sound Quality Jury Analysis: Paired Comparison and Semantic Differential. The former is a very straight forward test and can be administered to many test subjects, while the latter requires more concentration and must be used carefully [1].

During the Paired Comparison section of this study, jurors listened to pairs of sounds and asked which ones they preferred. The result of this test is a preference ranking of the sounds presented, but not a prediction of the preference score for untested sounds. For this, we need to identify specific objective sound metrics that are differentiators. This was done based on experience and the results of a Semantic Differential Test.

In the Semantic Differential Test, jurors are asked to rate a sound recording with regards to the descriptors mentioned above. For example jurors listened to a recording and were asked "on a scale of 1 to 5, how 'whiney' is this sound". Next, they listened to the same sound and were asked "on a scale of 1 to 5, how pleasant is this sound?" This was repeated with all descriptors and all recordings. The results of the Semantic Differential Test show us how people equate "whiney" and "annoying" in this example. These give us key insight into the objective sound metrics that affect subjective sound preference.

4.3 Analysis of Jury Results

The merging of the subjective results of the jury with the objective sound metrics was accomplished using a multiple linear regression analysis. Several potential metrics for each characteristic of the sounds were considered. Each was analyzed for correlation to the Juror's rankings of the sound and combined until, ultimately, a three-factor preference equation was developed that correlated very closely (adjusted $R^2=0.982$) with the results of the Paired Comparison rankings. Each factor within the preference equation represented a unique acoustic feature that affected the subjective preference of the jurors, and ultimately the preference of the end user. See Figure 3.



Based upon these results, subjective preferences can now be predicted for any sound by calculating Sound Quality metrics and their associated coefficients that were determined from multiple linear regression. One limitation of this method is that any sounds used in this fashion must be recorded in a similar manner (transducer placement and operating condition).

The sound preference equation developed via jury studies and sound quality analysis provided a repeatable, objective predictor of customer sound preference. Based upon the results of this study, production products as well as those within the development cycle were scored using the preference equation. This provided the ability to develop and recommend product design changes that could be specifically related back to an expected improvement to subjective preference by the end customer. However, as this information was used to improve the product's performance, the preference equation began to produce merits that were greater than 10 on a 0 – 10 scale. This implies that the sound of these improved products would be judged superiorly to those that were presented to jurors during the test. This prompted a decision to accept these values that were greater than 10, or to normalize the scale such that the best sound was scored as 10.

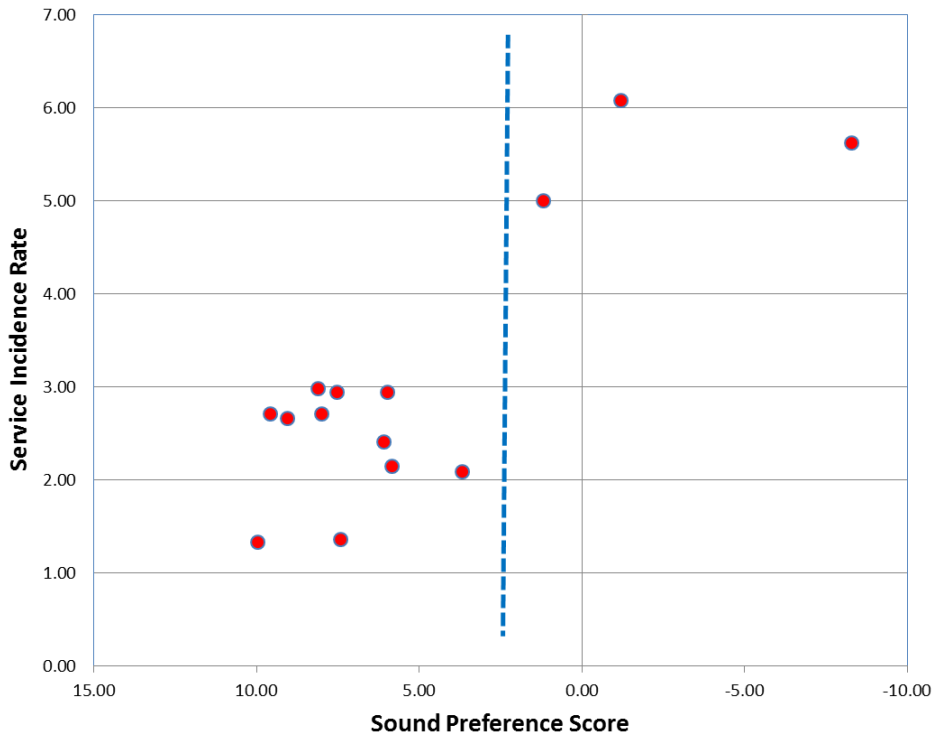
Both options carry considerations; if the merits greater than 10 are accepted, it can be seen as a source of confusion for Engineers who are accustomed to working with a 0-10 scale. If the scale is normalized, then all historical results would need to be adjusted also. Not only does this require significant labor, but this option would be a common point of confusion as well. For example, if part of a discussion includes the statement that a given product was rated as a “6”, it would need to be clarified if that “6” was on the original scale, or the revised scale.

After considering the options discussed above, it was decided that scores greater than 10 would be accepted. To confirm the validity of this method, an informal jury test was conducted to ensure that sounds which scored “15” were subjectively superior to those which scored “14”, “10”, etc. The merits were confirmed, and the preference equation continued to be used as it was originally developed.

5 DISCUSSION AND CONCLUSIONS

Once a database of preference scores on production units was established, the focus was shifted towards establishing preference score targets to utilize throughout development. A model-by-model comparison of the preference scores with the field incident rate of sound complaints yielded a clear and defined threshold. Models that scored above the threshold had very few, if any, sound related field issues.

The result of the correlation between sound preference scores and field service incident rates has allowed us to establish preference score targets for new product development teams that will give us confidence of customer satisfaction. See Figure 4.



Additionally the individual elements of the preference score algorithm provide more focused guidance to the product development engineers on areas to focus on to improve scores and ultimately customer satisfaction.

After successfully implementing this evaluation for standard operating conditions, a similar study was completed to expand this capability to transient test conditions such as start-up and shut-down noises as well as for steady-state operational noise.

6 ACKNOWLEDGEMENTS

The authors wish to thank the Management teams from Sub-Zero and Sound Answers for their direction and consideration of this paper.

7 REFERENCES

1. Meilgaard, Civille, Carr. 1991, *Sensory Evaluation Techniques – 2nd Edition*, CRC Press, New York, p.100
2. Cerrato, Gabriella. 2007, *Sound/Vibration Quality Engineering*, Sound and Vibration, www.sandv.com/downloads/0704jay.pdf
3. IEC 60704-1:2010, Household and similar electrical appliances - Test code for the determination of airborne acoustical noise