Sound Quality Studies of Front-Loading Washing Machines

David L. Bowen, Acentech Incorporated, Cambridge, Massachusetts

This article describes methods that were used for evaluating and designing sound quality improvements on a front-loading washing machine where the primary goals were to evaluate sound quality as determined by actual users and to measure the impact and value of various possible sound modification scenarios. These goals were accomplished in part by convening jury studies where jurors rated the sounds of various "virtual" and existing washing machines to identify and quantify how various components and mechanisms that operate within each cycle/mode of the washer affected consumer perceptions of sound quality expressed in terms of their ratings on the attributes "acceptability" and "overall impression."

As laundry centers have become more integrated into the main parts of the home, consumers and manufacturers have become more concerned about the sound and vibration generated by washing machines. Of particular concern are front-loading washers, particularly since they typically have a higher spin speed than top-loading machines, and their spin axis is horizontal. However, the sounds generated during the other cycles (fill, wash, drain, etc.) can also play a role in determining the overall "acceptability" of the washer and in forming an overall impression of washer performance, workmanship and brand image.

A manufacturer of a high-end front loading washing machine was concerned about these sound quality (SQ) issues and was interested in obtaining design guidance that could be applied to improve the SQ of its current and future models. This article describes the supporting investigations carried out to provide this design guidance.

An initial consumer jury study was conducted with the goal of determining which washer components and mechanisms have the greatest effect on sound quality (as expressed by jury ratings on two attributes, "acceptability" and "overall impression"), and the degree to which changes in these components would affect SQ. From this information, two alternative target designs emerged for improving SQ by similar amounts. Next, a second jury study was conducted to verify whether consumers would, in fact, prefer the sound of these two designs over the sound of the existing washer, or if one design was preferred over the other. This second jury also compared the targets and existing baseline sounds to the sounds of a prototype and a competitive front-loading unit. The resulting preference rankings enabled a more precise definition of the type of modifications that should actually be implemented on the existing washer.

Initial Jury Study

The goal of the initial jury study was to quantify how the various sound-producing components in the washer affect the two SQ attributes of acceptability and overall impression, as determined from a jury of washing machine users. This first required designing the jury test and then generating the component sounds that comprise the overall sounds heard during each cycle (fill, wash, drain and spin) of the machine, followed by the creation of a set of virtual washing machine sounds obtained by varying the component sounds according to a statistical design-of-experiments approach. This stimulus set was then presented to a consumer jury that ranked the sounds in terms of the two attributes. Regression relationships between the component variations and the resulting jury data were then developed, and these were used to select two candidate washer designs yielding similar degrees of improvement in terms of the attribute ratings predicted from the regression model. Each of these steps is further discussed below.

Design of Jury Test. The approach used for the initial jury test was to present to a panel of consumers the sounds of various virtual washing machines made up from extracted/simulated component sounds that were mixed together at different levels according to a statistical design-of-experiments (DOE) criteria. Jurors then noted their numerical ratings, on a scale of 0 to 100, for how acceptable and also for how well-built, reliable, and effective (overall impression) they thought these washers were, all based on their sounds while imagining the washers being located in the main part of their homes where they would hear them.

The basic idea for this type of SQ jury study is to break the total sound of the washer down into its various primary component sources for each of its operating modes, or cycles, so that an array of virtual washer sounds can be created by mixing together modified versions of these component sounds. The component sounds are varied in a quantifiable way over a range of values about their nominal (center) values, enabling us to explore subjective reactions within a space surrounding the existing baseline sounds for each operating mode of interest. The degree of component changes and the number of resulting mix combinations are governed by the particular DOE employed.

We chose to use a central composite DOE that we have had experience with in similar SQ jury studies on consumer products.[1-3] This design requires creating a number of sound mixes composed of various unique sounds, with each component varied by a certain number of steps, depending on the number of components to be varied, plus a number of repeats of the (unvaried) baseline sound. These steps can be any type of change that is quantifiable in some consistent way. In general, we want a single step to be a noticeable change, while not being so large that the largest number of steps called for in the design would put us out of the realm of the type of change that is actually feasible to implement. For this study, we chose to primarily vary the overall sound level of the components with a step size of 3.5 dB. In one case (the initial part of the fill cycle), we also varied the duration of the water-on and water-off events.

The central composite DOE employed was a circumscribed type, with the number of center points chosen so as to give uniform precision within the inference space. For our targeted jury size of 24 members, we needed to repeat each set of sounds three times in a randomized fashion.

Recordings. Generating the component sounds starts with recordings made of the product in a representative environment and distance from the unit during typical use. Since most of the noise from a large appliance like a washing machine appears in front of the user, monaural loudspeaker presentation is the most effective way to eventually present the sounds to jury members. While a true binaural recording and playback of an unaltered baseline sound may offer more realism to certain listeners, such realism is unlikely to be preserved once component sounds are edited and mixed together. It is unnecessary in any case if the sources are mainly ahead of the listener, since such a situation can be adequately simulated by a single loudspeaker.

For operating conditions, we typically used an 8-pound towel load with detergent, although we also made some recordings at twice this load. Warm water was supplied to the washer for all cycles, and the drain hose was put onto the edge of a plastic tub, to simulate drainage into a nearby laundry sink or dedicated drain. For the most part, we used the normal cycle and selected the extra-high spin speed. The washer was leveled during setup, and the degree of imbalance was monitored during spin. If the imbalance appeared too great (not representative of a typical spin), another recording session was initiated.

Component Identification and Extraction. The sounds during the various operating cycles of the front-loading washing machine were carefully evaluated to identify potentially important physical components and mechanisms that contributed to the overall sound for each cycle. The resulting components identified for varying are listed in Table 1.

The fill cycle was broken down into two parts to capture the short pulsating characteristics of this particular washer during the initial part of fill, and also to include the sound of a noticeable valve closure that occurs in the later fill portions when the fill/flow sound is also noticeably different than in the initial portion. To vary the water-on and -off times, we needed to present at least one cycle of water on, water off, water on to the jury. While the later fill portions are also pulsed, the durations of the water-on and -off segments were generally too long to include in a jury study of this type (8-13 seconds for water on and 11-18 seconds for water off, as opposed to 0.5-5 and 2-5 seconds, respectively, during the initial part of fill). So the duration of the complete water-on, water-off, water-on samples for fill 1 varied from 4.5 to 12.5 seconds. The durations of the sound samples for the other cycles were 4 to 4.5 seconds.

Generating the component sounds was accomplished primarily by applying to the basic recordings a combination of time-gating, filtering, and fade-in/out techniques. Spectral analysis, along with critical listening evaluations, were used to identify frequency values and ranges needed for the filter designs. For example, a low-pass filter set to 1100 Hz was used to extract the rotation-related sounds during steady-state, high-speed spin, followed by notch filtering to eliminate most of the motor-related sounds. Similarly, the windage noise during spin was extracted using a high-pass filter set to 1100 Hz along with notch filtering of various motor-related frequencies that had been previously identified. Care was taken to preserve phase relationships between component sounds in those cases where there were strong tonal characteristics. In one case, we made a modification to the washer operation itself to aid in extracting a component sound. This modification involved inserting a device into the wash drum designed to hold the clothes against the perimeter of the drum so that clothes "flopping" would be eliminated, leaving only the water slosh and motor sounds. High-pass filtering was then used to effectively attenuate the motor noise from the resulting recording while preserving most of the water slosh noise.

In addition to the sounds of the virtual washing machines made up from the components described above, the jury also heard the sounds of two additional washers – a prototype of the next model and a competitive unit. Appropriate sound samples from these units were chosen for use in most of the five cycles described above. While the sounds from these extra machines did not enter into the more extensive regression analysis performed on the jury response data for the varied component sounds of the current washer, their mean attribute ratings could nevertheless be compared against each other and to the baseline sound of the current washer.

Table 1. Components of various machine cycles identified for varying in initial jury study.

Fill 1 (Initial Part):
  A. Overall level of flow noise during "water-on"
  B. Duration of water-on portion
  C. Duration of water-off portion

Fill 2 (Later Part):
  A. Overall level of flow noise
  B. Overall level of valve closure noise

Wash/Rinse:
  A. Overall level of circulation pump noise during startup
  B. Overall level of circulation pump during steady-state operation
  C. Overall level of motor noise
  D. Overall level of "clothes flopping" noise
  E. Overall level of "water slosh" noise

Drain:
  A. Overall level of drain pump noise during startup
  B. Overall level of drain pump noise during steady-state operation
  (Water-related sounds associated with water exiting drain hose were also added for realism. But these sounds were not varied, since they are beyond control of washer.)

Spin (Steady-State, Highest-Speed Spin):
  A. Overall level of motor noise (primarily motor tones)
  B. Overall level of "rotation-related" (once-per-rev.) sounds
  C. Overall level of "windage" noise
  D. Overall level of high-pitched "whine" sound

Except for the wash cycle, where there were five components to vary, the cycles all used full factorial central-composite designs (CCDs); the wash cycle used a half-fraction CCD, so that we could obtain a reasonable number of sounds to present.[1]

Execution of Jury Test. A single 2-hour-long session with 26 jurors was conducted inside a fairly anechoic listening/conference room (same room as used for washer recordings). The jury consisted of 18 women and eight men, with seven jurors owning front-loading washing machines and 20 owning top-loaders (one juror owned both).

A single loudspeaker was located in the front of the room and was used to present the sounds to jurors seated at tables facing the speaker. The sound level was set so that the A-weighted levels at the juror positions were approximately what the jurors would experience if the washer were at the speaker location.

The jurors were first given verbal instructions followed by an example (practice) session using sounds. Jurors were instructed on the meanings of the two attributes, how to fill in the forms, and to imagine that the washer they were listening to was located where the speaker was. They were told to use a fixed interval scale of 0 to 100, with 0 being completely unacceptable and 100 being extremely acceptable for the acceptability attribute, and, for the overall impression attribute, 0 being a very negative impression and 100 being a very positive impression. A graphic representation of this scale was placed on the table in front of each jury member to remind them of the meanings of the rating value ranges.

Jurors were given 9 seconds after hearing each sound sample to write down their ratings for both the acceptability and impression attributes. The total number of sounds presented for each cycle, including the extra sounds from the prototype and competitor units ranged from 42 for fill 2 and drain to 105 for wash and spin. (The varying numbers are due to the different number of components varied for each cycle).

The jurors were also told that the washers they were listening to were all front-loading machines (brands not divulged). To assist the jurors in picturing front-loading washers, photographs of various models were placed on the walls around the room. At the end, jurors were given the opportunity to write down any general comments they may have had about the sounds.

Analysis of Jury Response Data. The data from all 26 jurors was entered into a computer and checked for accuracy and consistency. The data from two of the jurors was thrown out, since it appeared that for most of the cycles, these individuals were just writing down the same number for each sound.

The reduced sets of jury response data were then used as inputs to a response surface analysis configured to use a multiple linear regression to build a quadratic model of the relationship between the observed acceptability and impression ratings and the component sound level (or duration) changes. Basically this process tries to fit the entire set of individual juror ratings to a response surface model with all the linear, squared and two-way interaction terms included.[1] After eliminating a first round of observations with large standardized residuals (less than 5% of the total) and eliminating statistically insignificant coefficient terms, the analysis was re-run, resulting in a response surface model for each cycle that could be used to predict the effect on acceptability and impression of changes in the various components of the washer.

The equations resulting from the reduced-order regression analyses on the jury ratings are summarized in Table 2. These relate values predicted for acceptability and overall impression (on the 0-100 scale described previously) to changes in the component sounds from their as-recorded baseline conditions.

Associated with each of the equations in Table 2 are various statistical indicators arising out of the regression analysis that can be used to help assess the reliability and accuracy of the models.

One of the most important indicators is the regression coefficient R^2 which is the percentage of the observed variability that can be explained by the models described by the equations. We have typically seen R^2 values in the 25-55% range in many of the SQ jury studies we have conducted on consumer products. The values in this study ranged from about 10% for drain to 55% for fill 1, depending on the attribute. Some of the reasons for the relatively low R^2 values may be related to different internal scaling values that the individual jurors employed (e.g., different jurors using different subranges of the full 0-100 scale). Other than not using the data from the two jurors noted above, we did not explicitly try to account for any level, scaling or variability effects that may have been attributable to differences in the way individual jurors responded or made use of the scaling system.

It may also have been possible to obtain higher R^2 values either by introducing larger changes into how the components were varied or by changing some other yet-to-be-identified component. However, such changes have to be tempered with what is reasonable to achieve from an engineering point of view. The component noise levels for wash and spin, for example, were varied over a ±7 dB range (±5 dB for drain), which we felt was a reasonable amount that could actually be achieved. As noted above, we generally want a single step change to be a noticeable change while not being so large that the largest number of steps called for in the design would put us out of the realm of the type of change that is actually feasible.

In any case, the fact that the variability in juror judgments was large does not mean that the predictions of the effect of a change in component level are wrong, but simply that such a change may not be universally recognized as beneficial. The regression equations themselves are a tool that can be used to guide subsequent engineering efforts in the pursuit of increasing consumer satisfaction with the product. The verification study described below addressed some of the uncertainties discussed above by presenting the sounds of washers with supposedly higher sound quality along with the sound of the original washer to determine the degree to which consumers might prefer one machine over the other.

Table 2. Reduced-order regression equations obtained for sound quality attributes, "acceptability" and "overall impression."

Fill 1:
$\text{Acceptability} = 62.8 - 4.9\mathbf{A} + \mathbf{B} - 0.24\mathbf{A}^2 - 1.9\mathbf{B}^2$
$\text{Impression} = 64.6 - 3.5\mathbf{A} + 1.6\mathbf{B} - 0.15\mathbf{A}^2 - 1.8\mathbf{B}^2$
where: **A** = Change in overall level of water-on relative to baseline level (dB)
B = Change in duration of water-on relative to existing average duration (seconds)

Fill 2:
$\text{Acceptability} = 38.0 - 4.2\mathbf{A} - 0.46\mathbf{B} + 0.24\mathbf{A}^2 + 0.25\mathbf{B}^2$
$\text{Impression} = 45.0 - 3.3\mathbf{A} - 0.087\mathbf{B} + 0.20\mathbf{A}^2 + 0.13\mathbf{B}^2$
where: **A** = Change in overall level of flow noise relative to baseline level (dB)
B = Change in overall level of valve closure sound relative to baseline level (dB)

Wash:
$\text{Acceptability} = 56.9 - 0.63\mathbf{A} - 0.34\mathbf{B} - 0.24\mathbf{C} - 1.5\mathbf{D} - 0.09\mathbf{A}^2 + 0.05\mathbf{B}^2 - 0.051\mathbf{D}^2 + 0.13\mathbf{AB} + 0.069\mathbf{AD} - 0.098\mathbf{CD}$
$\text{Impression} = 57.7 - 0.54\mathbf{A} - 0.19\mathbf{B} - 1.7\mathbf{D} - 0.073\mathbf{A}^2 + 0.064\mathbf{B}^2 - 0.051\mathbf{D}^2 + 0.11\mathbf{AB} + 0.11\mathbf{AD} + 0.098\mathbf{BD}$
where: **A** = Change in overall level of circulation pump startup sound relative to baseline level (dB)
B = Change in overall level of circulation pump steady-state sound relative to baseline level (dB)
C = Change in overall level of motor noise relative to baseline level (dB)
D = Change in overall level of clothes "flopping" sound relative to baseline level (dB)

Drain:
$\text{Acceptability} = 48.0 - 0.89\mathbf{A} - 1.4\mathbf{B} + 0.29\mathbf{AB}$
$\text{Impression} = 50.8 - 0.74\mathbf{A} - 1.1\mathbf{B} + 0.27\mathbf{AB}$
where: **A** = Change in overall level of drain pump startup sound relative to baseline level (dB)
B = Change in overall level of drain pump steady-state sound relative to baseline level (dB)

Spin:
$\text{Acceptability} = 50.2 - 0.51A - 2.4\mathbf{B} - 1.5C - 0.068\mathbf{A}^2$
$\text{Impression} = 53.5 - 0.43A - 1.9\mathbf{B} - 1.5C - 0.067\mathbf{A}^2 + 0.11\mathbf{BC}$
where: **A** = Change in overall level of motor sound relative to baseline level (dB)
B = Change in overall level of rotation-related sounds relative to baseline level (dB)
C = Change in overall level of "windage" noise relative to baseline level (dB)

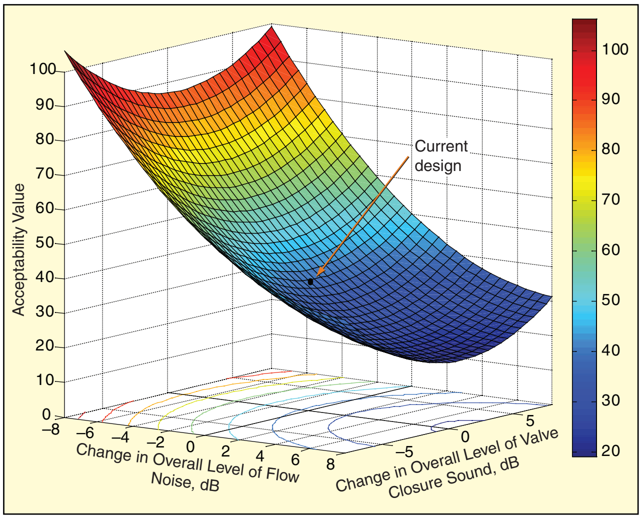


Figure 1. Regression model response surface for acceptability of later parts of fill (fill 2), showing how value of this attribute changes as sound levels of flow and valve closure noise change from present values.

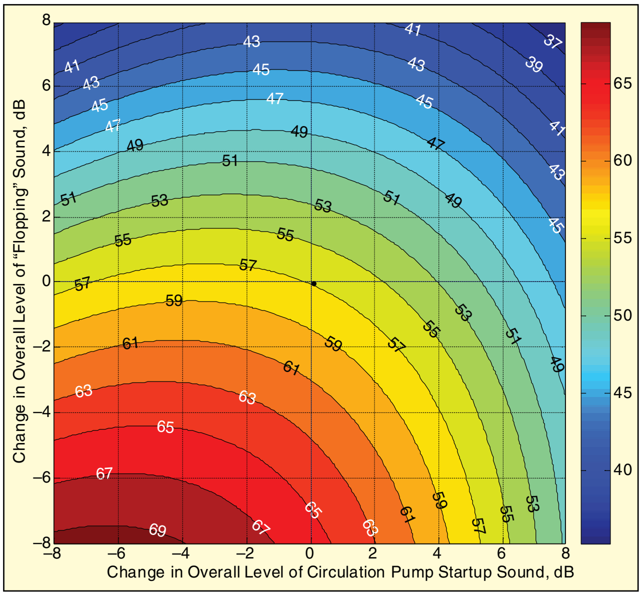


Figure 2. Contour plot resulting from regression model of jury responses to the wash cycle sound, showing dependence of acceptability on changes in levels of circulation pump and clothes-flopping sounds; circulation pump steady state and motor sounds held fixed at 0 dB change from present levels.

To facilitate interpretation of the relationships described by the regression equations in Table 2, response surface plots of the models can be produced. Such plots can be useful for quickly gauging how sensitive the attribute ratings are to changes in the various components and for determining the most efficient way to improve the ratings by a certain amount. Figure 1, for example, shows the response surface generated for acceptability of the sound of the fill-2 cycle, as the component levels are varied over a ±8 dB range around their present level indicated by the black dot. (Exploring the space with increased sound levels is included, since we have sometimes seen SQ attributes increase with increasing levels). In this case, since there are only two components (levels of flow noise and valve closure sound) that were varied, a three-dimensional plot can be used to convey the entire behavior of the model. Figure 1 shows that the overall level of flow noise has a much larger influence on this attribute than does the overall level of valve closure sound, as would be expected given the ratio of coefficient values

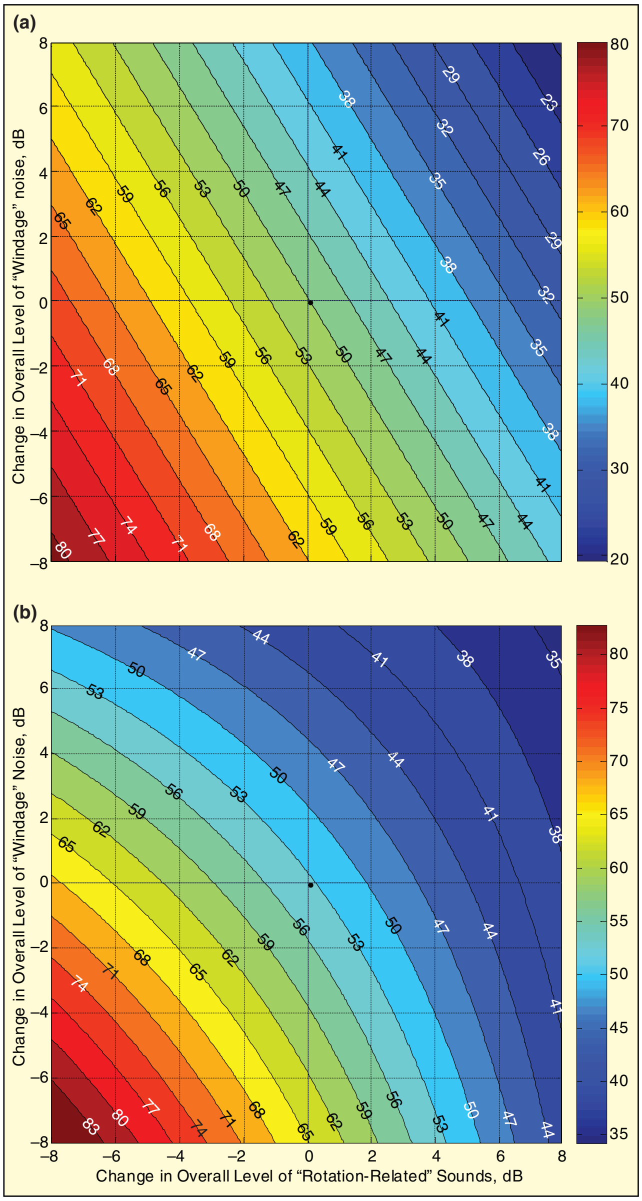


Figure 3. Contour plot resulting from regression model of jury responses to steady-state portion of spin cycle sound, showing dependence of (a) acceptability and (b) impression on changes in levels of rotation-related and windage sounds; motor sound held fixed at 0 dB change from present level.

for these components in the corresponding regression equation.

Visualizing the models for the other machine cycles, where there are more than two components, is not as easy, since there are more than three dimensions. However, we can look at contours or slices through the response surfaces generated by varying two components at a time while holding the remaining components fixed at some value such as 0 dB change from their baseline values. Examples of such contour plots are shown in Figure 2 for acceptability of the wash cycle, and in Figures 3a and 3b for acceptability and overall impression of the spin cycle. For wash and spin, the plots shown correspond to varying only those components with the largest coefficients in the regression equations (the motor noise level had little influence on the ratings for these cycles compared with the other components). The intersection of the solid lines in each contour plot represents the current baseline condition. Keep in mind that since the model essentially results from a curve-fitting process, the associated confidence in the model generally decreases the closer one gets to the boundaries of the variable space explored (e.g., near and beyond the edges and corners of the plots).

The contour plots and regression equations can also be used as guides to help pick out candidate target designs for improving the attribute ratings by a fixed number of points. For example, we could achieve a 12-point increase in the acceptability rating of fill 2 by decreasing only the flow noise by a little over 2 dB, or by decreasing only the valve closure noise by 6 dB, or by some combination of the two that ends up on the same contour line. Going through this type of analysis for acceptability (impression was generally closely correlated with acceptability in this study), supplemented by use of the regression equations for wash, suggested the following two preliminary target designs for improving the ratings by approximately 12 points:

Preliminary Target Design 1:

- Decrease flow noise during fill 1 by 6 dB.
- Decrease flow noise during fill 2 by 2 dB.
- Decrease circulation pump and clothes flop noise during wash by 4 dB.
- Decrease drain pump noise during drain by 3 dB.
- Decrease rotation noise during spin by 4 dB and windage noise by 2 dB.

Preliminary Target Design 2:

- Decrease flow noise during fill 1 by 6 dB.
- Decrease valve closure noise during fill 2 by 6 dB.
- Decrease startup noise of circulation pump during wash by 8 dB and its steady-state noise by 7 dB.
- Decrease startup noise of drain pump during drain by 6 dB and its steady-state noise by 2 dB.
- Decrease rotation noise during spin by 2 dB and windage noise by 6 dB.

The corresponding sounds for these or other designs can then be created and presented to a jury in a paired-comparison test (along with the unaltered baseline sounds) to determine whether there is a clear preference for one design over the other. This was a focus of the paired-comparison study described next.

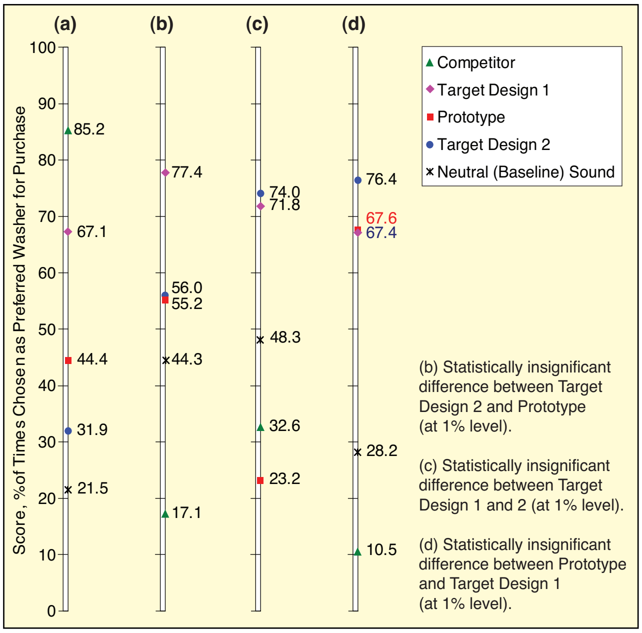
Follow-Up Verification Jury Study

Next, a study was designed where we sought to compare the sounds of the two target washers identified above to each other, and also to the sounds of the original (unchanged) washer, the competitive model, and the new prototype. The two candidate designs offered different ways to achieve similar degrees of improvements in sound quality as conveyed by the acceptability and overall impression attribute ratings predicted using the regression relationships developed from the results of the first jury study.

The paired-comparison (forced-choice) technique is often used to determine if there is agreement among consumers that some versions or brands of a similar product are perceived as being better (preferred) than others and, if so, how they rate against each other. The primary purpose of the verification study described here was to determine whether there is strong consumer preference for the sound of one design over another so that a final design could be selected for implementation. Preference rankings of the different washer sounds are obtained, and these results are used to quantify how close or far apart the washers are to each other in terms of consumer preference.

Structure of Paired-Comparison Jury Test. The sounds of five different washers in each of four main operating modes (fill, wash, drain and steady-state highest-speed spin), were presented to a consumer jury one pair at a time. In addition to the two target designs described, the washer sounds consisted of the baseline (neutral) sound of the current front-loading washer as created by mixing together the unaltered component sounds, the sound of the competing model, and the sound of a prototype washer.

The two target designs used for the paired-comparison study were slightly modified versions of the preliminary target designs described above, with the fill 1 cycle eliminated, since its short pulsating fill sounds are not easy to compare without confusing jurors (and the competitor unit did not have this type of fill in the first place). The sound for wash was taken about 0.75 second before the start of a tumble and did not include the circulation pump startup sound (but it did include the steady-state sound of the circulation pump). This was fairly representative of the typical sound heard at the beginning of a wash tumble. The high frequency whine sound during spin was not included, since this component was found to have a fairly small influence on the ratings during the first jury test.



(b) Statistically insignificant difference between Target Design 2 and Prototype (at 1% level).

(c) Statistically insignificant difference between Target Design 1 and 2 (at 1% level).

(d) Statistically insignificant difference between Prototype and Target Design 1 (at 1% level).

Figure 4. Preference rankings for purchase based on sounds of (a) fill, (b) wash, (c) drain and (d) spin cycles (95% confidence bounds = ±5 on values).

As with the first jury study involving the fixed-interval scale, the design of a paired-comparison study involves trade-offs between session length, the number of jurors to recruit and the desired confidence of the results. Since there were five sounds being compared, there were 10 unique pairs to present for each of the four washing machine cycles being considered. For a paired-comparison study, either the number of jurors or the number of repeated presentations needs to be fairly high to gain reasonable confidence in the results. We typically desire about 100 judgments for each of these unique pairs and, to eliminate possible bias errors, another 100 judgments to be made on the corresponding reverse-ordered pairs. For a 95% confidence level, the resulting error rate associated with these 200 judgments should be 6.7% if all the judgments were truly independent (that is, produced by 200 different jurors). We can relax this requirement somewhat by repeating the pairs to a smaller group of jurors so that we still have a total of 200 judgments, with the resulting error rates lying somewhere between that expected for the smaller number of jurors and what would be expected if we actually had 200 jurors. Due to project budget and other considerations, we targeted 60 jurors instead of 200, implying an error rate somewhere between 6.7 and 12.2%. This strategy enabled us to hold two 1-hour sessions instead of 8 sessions, with the jurors hearing all pairs four times (two in the forward direction and two in reverse).

Executing Paired-Comparison Test. Recordings of the prototype and competitor washers were made under the same conditions as used for the original baseline washer recordings. After selecting and extracting representative sound samples from these recordings, and creating the sounds of the two target designs (and comparable neutral/baseline sounds) for each of the cycles considered, sound pairs were then created for presentation to the jury. Each sound in a pair was 3 to 4 seconds long, depending on the cycle, with a pause of 0.5 second between the two sounds (except for wash, where a longer pause of 1 second was used). After randomizing the presentation order of the pairs for each cycle (over all pair possibilities and all repeated runs), a pause of 5 seconds was inserted between each pair to give the jurors time to make their choice.

Before the sound pairs were presented, the jurors were instructed to choose the front-loading washer they would buy based only on its sound, assuming all other features, performance, etc. were comparable between the two washers. The preference question was cast in terms of purchase preference, since this is a direct indicator of consumer acceptance of the washer sound, encompassing judgments on all the various sound quality attributes that might go into making a purchase decision. Purchase preference is also more suited to a forced-choice type of jury test than it (or its scalable version, purchase likelihood) is to a ranking type of test such as the first jury test. As with the first test, a single loudspeaker was used for playback of the sounds. 30% of the 59 jurors that showed up for test owned front-loading machines.

Analysis of Paired-Comparison Results. The response data from the jurors was entered into a computer and checked for accuracy and consistency. All of the data from every juror was used in the subsequent analysis for obtaining the overall purchase preference rankings of the five washers.

To do this, we first determine the total number of times a given washer was chosen across all jurors and all presentations of that sound. Then this rank sum is normalized by the total number of times that washer could have been chosen if it were always chosen in every instance in which it was presented. Figures 4a-d display the resulting purchase preference rankings expressed as a percentage of times a particular washer was chosen for each of the four cycles considered in the paired-comparison study.

Given the number of sounds presented, the number of jurors and the number of repeated pairs, there is a certain confidence interval associated with these rankings. In this case, the 95% confidence bounds are ±5 around each of the score values shown in Figure 4. Also, we find that differences of less than 6.6% are statistically insignificant at the 1% level. At this indicated difference level, for example, there is only a 1% chance that both washers are, in fact, equally preferred by the jury. While such small differences occurred once each for the wash, drain and spin cycles, we see that most of the preference differences are significant. We note that there are noticeable differences between the sounds of the baseline and target designs that caused the consumer jury to prefer the target designs over the baseline unit. The degree to which they preferred one washer over the other varies by cycle and is indicated in Figure 4 by how far apart the score values are for each washer.

The recordings made of the two extra units (the prototype and the competitor) for the paired-comparison study naturally had operational conditions that differed more than they did between the more controlled cases of the two made up target designs and the unmodified (neutral) washer sound from which the target sounds were derived. For the drain cycle, it is possible that the jurors may have focused on differences that existed between the water splashing/dripping sounds in the recordings made for the prototype, competitor and target/baseline group. This is likely the reason why, in this cycle only, the prototype scored lower than the baseline washer.

We also note that the sound of the competitor washer was the least preferred during the wash and spin cycles. The competitor recording for wash appears to have a louder clothes-flopping component in its sound compared with both the prototype and baseline units. During spin, the competitor recording seems to have more high-frequency windage noise than the prototype and baseline units. This latter observation appears to be consistent with our finding that an effective way to improve the SQ for spin is to decrease the windage noise level.

Selection of Final Target Design

The results from the paired-comparison verification study indicated that except for the fill cycle, at least one of the two target designs was the most preferred washer sound. Furthermore, the degree of preference for the top target design over the next highest ranked washer was fairly substantial. While the prototype did appear to offer substantial sound quality improvement over the baseline washer, the target designs offer a way to improve SQ to an even greater extent.

For the fill and wash cycles, target design 1 was preferred over target design 2; in the drain cycle they were about equally preferred. For spin, target design 2 was slightly preferred over target design 1 and the prototype. Based on these findings, we recommended implementing changes that approximate those suggested by target design 1 for the fill, wash and drain cycles, and by target design 2 for the spin cycle. Decreasing the higher-frequency windage noise more than the lower-frequency rotation spin noise, as suggested by target design 2, is likely to be easier to implement than the opposite strategy suggested by target design 1. The resulting final target design can be summarized:

Final Target Design:

- Decrease flow noise during fill by 2.5 dB.
- Decrease circulation pump and clothes-flopping noise during wash by 4 dB.
- Decrease drain pump noise (both startup and steady state) during drain by 3.5 dB.
- Decrease rotation noise during spin by 2 dB and windage noise by 6 dB.

Conclusions

An initial jury study was used to obtain ratings on the sounds of a number of virtual washing machines created by combining component sounds at different overall levels to identify and quantify how the various components and mechanisms that operate within each cycle of the washer affect consumer perceptions of sound quality. For the particular example highlighted, sound quality was expressed in terms of consumer ratings on two attributes: acceptability and overall impression. Regression models between the resulting SQ attribute ratings and sound level changes of the various washing machine components were developed and then used to identify modified candidate designs that would increase the attribute ratings by a fixed amount (12 points out of a 0-100 scale). The sounds of two of these target designs were next presented to a second jury in paired-comparison fashion to determine purchase preference rankings when compared to each other and to the sounds of the existing washer, an existing prototype, and a competitor's washer.

We found that both target designs were usually the most preferred when compared with the other three washers and that jurors had a definite preference for the target designs over the baseline washer. Based on these results, a final composite target design consisting of reducing the overall noise levels of six of the components by 2 to 6 dB was selected for implementation. Of course, the success of the final design will depend on the degree to which such reductions can actually be achieved. But since the method by design focuses on feasible engineering changes that can be made to individual physical components and mechanisms, it is reasonable to expect that such modifications can be realized. (If they cannot be, then the regression models offer a way to determine the impact of lesser changes on sound quality).

For a device with different operating modes like a washing machine, it may sometimes be desirable to determine an overall SQ rating. One possible way to achieve this might be to weight the various SQ ratings obtained for each machine cycle according to the percentage of time the machine spends in each mode. However, the utility of this approach diminishes if there are many different combinations of modes (such as the various cycles available in most modern washing machines). It is also unknown whether such a simplistic approach would adequately capture situations where one particular cycle rates especially low.

A disadvantage of the method described is the obvious need to conduct consumer jury tests. But once data are obtained from such tests, the regression models that result can be applied to evaluate SQ on similar products with comparable components and mechanisms. For example, we should be able to use the models on other front-loading washers but probably not top-loading washers and definitely not dryers. We can also form relationships between the jury ratings and various calculable sound quality metrics to aid in predicting consumer response.[4]

References

1. Myers, R. and Montgomery, D., *Response Surface Methodology: Process and Product Optimization using Designed Experiments*, John Wiley & Sons, New York, 1995.
2. Lyon, R. H. and Bowen, D. L., "Designing Quiet Products," *The Bridge*, National Academy of Engineering, Vol. 37, No. 3, pp. 11-17, 2007.
3. Lyon, R. H., *Designing for Product Sound Quality*, Marcel Dekker, New York, 2000.
4. Bowen, D. L., "Correlating Sound Quality Metrics and Jury Ratings," *Sound & Vibration*, pp. 12-15, September 2008.

The author can be reached at: dbowen@acentech.com.