Appliance Engineer, May 2001

Use of Consumer Listening Panels to Enhance Sound Quality

David L. Bowen, Corporate Consultant, RH Lyon Corp, Cambridge, Mass.
James P. Carow, Senior Engineer, Whirlpool Corporation, Benton Harbor, Mich.

Sound can contribute to a consumer's overall evaluation of a product in terms of its acceptability and function. The need for manufacturers to develop high-end and international markets has made product sound an important product attribute.

Achieving a desired sound shares some similarities with achieving the right color, surface texture or shape for the product, in that sound involves peoples' reactions, and the more favorable sound can be determined to only a very limited degree by physical measurements. However, the latter aspects are generally independent of other aspects of the design, while sound is affected by many of the basic features of the product. The sources of motion (motors, actuators, solenoids, etc.), their interconnections, and the structures that hold them in place all participate in determining the sound that is generated. The design has to somehow connect the choices for these physical components to the perceptual reactions of customers.

Manufacturers are often faced with negative reactions to the sound of their products. However, sound may enhance as well as detract from the pleasure in using a product, and may also indicate how well the product is working. This article discusses procedures that can be incorporated into the design process so that the positive attributes of the sound are identified and enhanced, and the negative ones are reduced, using the Whirlpool-designed "Calypso" washing machine as an example. The approach used combines product sound analysis to identify the major components responsible for the overall noise, along with jury testing of a virtual array of different possible designs. This process can provide specific design goals for the sound of a product in terms that product planners can specify (such as likelihood of purchase, perceived effectiveness, overall acceptability, etc.) and engineers can deal with (a reduction in water splash noise or a change in motor sound, for example).

Identifying Sound Quality as an Issue

The Calypso Wash Motion washer is a new technology, high efficiency vertical axis washing machine. Energy efficiency is achieved by replacing the typical deep fill of water into the tub with a recirculating shower, and by driving the machine with a variable voltage, current and frequency motor. The traditional agitator has been replaced with a "washplate" which wobbles (nutates) about a central pivot, producing an up and down motion of the clothes. With these design innovations came new washing machine sounds. Broadband high frequency sound was produced by splash and water interacting with the motion of the washplate and clothes. New techniques for controlling the drive motor produced motor sounds unlike the hum that consumers are accustomed to.

Early focus group studies, using development prototype units, indicated that consumers were concerned not just about the overall sound level of the washer (this was more of a concern for those who had their laundry centers located in places other than a basement or garage), but also about the "nature" of the sounds during all machine cycles. The sounds were such that some participants thought there could be something "wrong" with the washer (thus leading to a high number of service calls), or that it sounded so "powerful" they thought clothes could be damaged. Although these impressions did not correlate with reality (in fact, the Calypso wash system helps reduce wear on clothes), the focus groups clearly indicated that the washer sound was currently unacceptable, and needed substantial improvement before such a product could be introduced to the market.

While the focus group studies were a useful and effective means for determining the relative importance of sound in general, and for discovering the nature of the "noise problem", they did not yield quantitative information that could be used to provide design guidelines for improving the sound. In order to get such information, a series of "jury tests" using consumer listening panels was designed and carried out.

Measuring the Perception of Sound via Consumer Listening Panels

The focus group studies identified various subjective impressions that consumers felt were important, and which were conveyed by the washer sounds. These impressions, or perceptual attributes, included Loudness or "strength" of the sound, Gentleness of the washer, machine Effectiveness, perceived product Quality, and overall Acceptability of the sound. Of these, only Loudness has been shown to be reliably predicted using a physical measurement (e.g., Stevens or Zwicker Loudness can be calculated based only on a microphone signal). Other so-called sound quality metrics have been developed for one attribute of a single product (e.g., jet aircraft), or for no product at all, having been developed using simple sound stimuli such as tones, clicks, and bands of noise which, by design, convey no connotative meaning. However, such meaning cannot be avoided in the application of psychoacoustics to product sounds. The focus of psychoacoustics has been on how the auditory system assembles and processes acoustic information, peripherally and centrally. Traditional psychoacoustic methods do not, therefore, offer guidance for predicting listener responses to the more complex, actual sounds of products.

In order to determine the relationships between possible design changes in the washer and consumer responses in the various attribute categories identified above, special sound quality jury tests were designed. These tests were used to quantify how sensitive consumers were to changes in the sounds of specific components and mechanisms within the washer (in the context of the overall machine sound – jurors do not listen to individual component sounds), so that it could then be determined which modifications would maximize the "sound quality", as reflected by higher rating scores for the attributes. For example, while sounds such as tones or transient clicks may not significantly affect the overall A-weighted sound level, they can nevertheless be audible and thus potentially important in affecting the overall "acceptability" of the washer. In these jury tests, consumers listened to sound samples of several "virtual" washing machines, and then rated the attributes as the various component sounds were changed. In order to preserve the spatial relationship between a listener and an actual washing machine, the sounds were presented through loudspeakers rather than over headphones.

Generally, the way and the degree to which component sounds are changed is governed by those components identified as being potentially important, and by the statistical design of the tests. For the washing machine, there was interest in evaluating the sound quality in each of the four major machine cycles (fill, wash, drain, spin), and in each cycle there were different contributors to the overall sound. Analysis of the sounds from an early prototype, along with feasibility considerations, resulted in the following initial breakdown for the component sounds in each cycle:

- Fill cycle: transient click as fill valve opens, flow noise through valve, water dribble/splash.
- Wash cycle: motor, drive-train (gears), water splash sounds generated by washplate nutation, pump in recirculation mode.
- Drain cycle: motor, drive-train, washplate generated sounds, pump in drain mode.
- Spin cycle: click at start of cycle, whine during speed-up, motor hum, basket "rotation sounds", windage/water sounds, spray rinse noise, click at end of cycle.

Due to its length and complexity, the spin cycle was separated into an initial "speeding-up" portion, and a steady-state portion. Except for the windage/water sounds during spin, all of the component sounds listed above were considered to be potentially important and feasible to change, and were thus earmarked for varying. In the initial jury tests, these variations took the form of overall amplitude modifications (such modifications are usually sufficient for assessing each component's relative importance to the attribute ratings). In some of the subsequent jury tests, however, variations in the motor sound were tied to variations in the motor supply waveform, resulting in different degrees of harmonic content or "buzziness". Other jury tests utilized a frequency dependent attenuation of the wash cycle sound, designed to simulate the effect of varying amounts of sound attenuation as the wash rate and load size were also varied.

Sample recordings of the component sources next had to be generated. To accomplish this, either each component of interest was operated by itself, or each component sound was "extracted" from recordings made while the entire or partial unit was operating. All of the basic washing machine recordings were made in a hemi-anechoic room at the manufacturer. Some of the component sources such as the pump, flow through the fill valve, and the transient "clicks" were relatively straightforward to operate independently without other sounds interfering, and so could be recorded directly. However, since the motor sound was highly dependent on load, recordings of the motor operating by itself (achieved by removing its drive belt) could not be directly used to provide the sound of this component. Instead, the motor tones in these "unloaded" recordings were adjusted to match the motor tones present during normal washer operation.

Most of the other component sounds, including the motor sounds in some of the other cycles, had to be extracted from composite recordings by using a combination of time windowing and filtering. Bandpass, band-reject and notch filtering techniques were utilized to isolate each sound of interest, once the principal frequency characteristics of the component sound had been identified through spectrum analysis. To assist in such analysis, and to provide a tracking signal for notch filtering, a motor tachometer signal was also recorded simultaneous with all microphone signals. This tachometer signal was also used to perform slight "speed adjustments" in the component recordings before mixing them back together, in order to compensate for small variations in rotational speeds from one recording to another.

The component sounds in each cycle were next mixed together at various relative amplitudes, in order to create a large number of “virtual” washing machine sounds. Both increases and decreases in the amplitudes were used. The number of variations, the changes in component sound levels, the number and sequencing of the presentations to jurors, and the number of jurors were determined from statistical design-of-experiments criteria, for the given number of components to be varied. A “central composite” design was used, which was structured to recover from the jury data a bilinear, quadratic regression function relating changes in the component sounds to the perceptual attributes. This design is particularly efficient, which is advantageous in minimizing the number of different sounds to present, in order to avoid possible listener fatigue.

Once the set of sample sounds were prepared, the jury members (recruited according to marketing criteria) were instructed in a scaling method and in the meaning of the attribute judgments they were to make (i.e., the meaning of “Effectiveness”, “Acceptability”, etc.). These jury tests employed a magnitude estimation method of numerically rating the sounds on each attribute, in which participants make up their own "internal scale" after receiving training on the methodology. Not restricting the jurors to a fixed interval such as "1 to 10" avoids the possibility of crowding near either end of the scale (e.g., the "Olympics Syndrome"). At the end of the listening sessions, jurors gave their ratings corresponding to certain benchmark descriptors such as Very Acceptable, Moderately Acceptable, etc. In this way, the numbers entered by all listeners could then be renormalized to the same scale using a piecewise linearization scheme. A scale of 0-100 was chosen for this, where, for example, "0" would correspond to "Very *Un*acceptable" and "100" to "Very Acceptable". Jurors were typically exposed to each sound sample for 5 seconds, followed by 5 seconds of silence for providing their ratings.

Interpreting the Responses of the Jury

After checking the jury responses for statistical reliability (including confirmation that the Loudness ratings were well-correlated with independent measurements of overall A-weighted sound levels), the results were analyzed to determine regression equations relating changes in the component sounds from their baseline values to changes in attribute ratings for each of the four major machine cycles. Typically, such regression equations account for about 35% to 50% of the observed variability in these types of jury tests with sounds. Example equations for the Perceived Quality and Gentleness attributes obtained for the wash cycle sound of a second generation prototype (in which the gear noise had been eliminated by replacing the gear box with a redesigned drive train) are given below, where statistically insignificant terms have been dropped:

$$\textit{Quality} = 48 - 5.9\mathbf{a} - 1.8\mathbf{b} - 3.6\mathbf{c} - 0.37\mathbf{d} - 1.8\mathbf{a}^2 - 0.21\mathbf{b}^2 - 0.2\mathbf{c}^2 - 0.14\mathbf{d}^2 - 1.2\mathbf{ab} + 1.5\mathbf{ac} + 0.3\mathbf{bc}$$
$$\textit{Gentleness} = 43 - 2.8\mathbf{a} - 1.5\mathbf{b} - 5.3\mathbf{c} - 0.14\mathbf{b}^2 + 0.19\mathbf{bc}$$

where **a** = index denoting motor waveform "buzziness", in step number (-2 to +2)
b = change in motor sound level, dB
c = change in wash plate sound level, dB
d = change in recirculation pump sound level, dB

Such relationships represent response surfaces that, when examined together, can be used to determine how to change the components so as to optimally increase the attribute ratings. The regression functions can also be used as predictor equations to give guidance in answering such questions as "how much reduction of the motor noise or the water splash noise associated with washplate motion is 'enough'?", or in predicting the change in each attribute associated with anticipated changes in component sounds. It must be kept in mind, however, that the relationships may not be valid outside of the variable space actually presented to the jurors, which, in this case, was generally a range of ±7 dB from the baseline component sound levels.

By holding selected variables (e.g., the changes in component sounds) fixed, these relationships can be visualized through the use of contour diagrams, such as those presented in Figures 1 and 2 for the perceived Quality and Gentleness functions shown above. Fig. 1 shows the regression function for Quality as a function of motor and pump noise levels, with the washplate level and motor waveform "type" held fixed at their baseline values. This function indicates a maximum of perceived Quality in the wash cycle if the motor sound level were reduced by about 4 to 5 dB (from its baseline value indicated by the point at (0,0) in the middle of the plot), although the improvement in this particular rating is slight. The function also shows that the pump noise has much less of an influence on this attribute than does the motor noise level.

In general, the regression functions may not show a simple maximum as indicated in Figure 1, but there will usually be a gradient that will indicate the direction for favorable change. Such a situation is illustrated in Figure 2, which shows the Gentleness rating for the wash cycle as a function of washplate and motor sound levels. In this figure, the pump level is held fixed at its baseline value, but the motor waveform type has been set to a less buzzy value of "-1". The results in Fig. 2 indicated that, unless the motor sound level were to substantially increase, the most efficient way to improve the Quality rating for the wash cycle would be to primarily decrease the washplate-related sound, along with a slight decrease in motor sound. A similar analysis for the spin cycle, however, suggested a larger decrease in motor noise was needed in order to substantially improve the attribute ratings in that cycle.

These types of evaluations were carried out on all of the attribute regression functions for each machine cycle, with the objective of determining which component changes would maximize the Gentleness, Quality and Effectiveness ratings while minimizing the Loudness ratings. An emphasis was given to the wash cycle, proportionate to the larger amount of time the washer spends in this cycle. To assist in this effort, an iterative optimization routine was run on the regression functions, with the objective of minimizing Loudness, given minimum acceptable values for the other three attributes, and bounds on the component changes. Several candidate "designs" emerged from this process, all of which yielded similar improvements in the attribute ratings (conceptually, it is easy to see how this can be so, given the infinite number of possible designs along a single contour line of equal attribute rating).

Further Improvements

Although the candidate designs that result from the types of evaluations described above may all predict a similar degree of attribute improvement, they might be quite different to implement in terms of cost, availability of components, or compatibility with the parts distribution system. These possible designs can thus be regarded as competing between themselves, in addition to competing with products already in the marketplace. To address such issues, it is often desirable to carry out a supplemental "verification" jury test in order to compare the sound of the candidate designs to each other and to similar, existing models. In such a test, the candidate designs are presented to jurors using the method of paired comparisons, where the selection criteria may be something like “purchase preference”. This type of test has some advantages over the magnitude estimation test described earlier for design improvement, but it requires many more respondents.

Such a Purchase Preference test was carried out on two of the early candidate designs that emerged from the jury results described above (these designs differed primarily in their motor waveform types). The two designs were compared to each other and to the sounds of five other, existing washing machines in the marketplace. The results showed that consumers had about an equal purchase preference for the two designs in the wash cycle, and a slight, but statistically meaningful, preference for one over the other in the spin cycle. However, the results also showed that these two designs were approximately in the "middle of the pack" in terms of the five other models. A subsequent paired-comparison evaluation using the sound of an updated prototype unit and preferences expressed in terms of the Quality, Effectiveness and Gentleness attributes confirmed that there was room for further design improvements, primarily in the sound of the wash cycle. To assist in this effort, a final magnitude estimation jury test was carried out in which the wash rate, wash load and sound level were varied, with jurors making judgments on the three attributes above plus overall Acceptability. This time, the sound level was varied according to frequency dependent attenuations that were designed to simulate the effect of varying amounts of noise control treatments applied to the washer.

The results from this last jury test were used in selecting an appropriate wash rate for each load type, determining final wash cycle loudness specifications, and designing the final noise control treatment configuration that would achieve these specifications. This treatment included pads on the front and side panels of the cabinet, fiberglass wrapping around the wash tub, motor vibration isolation, and a molded component under the lid along with a lid seal to contain washplate splash noise. The optimum waveform for the motor was selected according to one of the prior jury tests.

Conclusions and Future Research

Product designers make choices regarding structure, materials and components in a product. Among the tools they need are ones that allow them to anticipate the effect of those choices on the perceived sound quality of the product, and not just on the resulting overall A-weighted level or loudness. One such tool is the special form of sound jury testing described in this article, where jurors listen to an array of virtual products created by altering the sounds of the major components in the product. Results from such tests are analyzed in order to determine regression equations between ratings provided by the jurors on various subjective attributes identified as being important, and changes to the component sounds. Information is then available to optimally guide the subsequent acoustical engineering effort, both in terms of identifying and prioritizing components in need of modification, and in quantifying the amount of component alterations needed in order to appreciably increase the attribute ratings. Application of these methods during the development of the Calypso washing machine helped set design goals, and provided input for design changes that substantially improved the sound from a consumer point of view.

The procedure described in this article result in designs that start with the current configuration of a product, and make changes that improve its perception. On that basis, the procedure is acceptable and useful. But it does not answer the question of whether there is a "best" design for sound, perhaps removed from the present design by some distance. In making incremental changes to the present design, the opportunity for a big improvement in perception may be missed by ignoring a big "hill" not too far away. To address issues such as this, RH Lyon Corp is currently conducting research, supported by the National Science Foundation, into a new approach to setting goals for designing product sound. One outcome of this effort has been the development and use of "sensory profiles" for sound as an intermediary between measurement metrics and user perception. Manufacturers of products from wines to perfumes to foods use sensory profiles to set goals for the perception of those products. The research has shown that it is possible to produce favorable acoustical sensory profiles for products, and to use those profiles to predict customer preferences. Such sensory profiles then have the potential to act as a bridge between customer perceptions in the marketplace and design decisions in the engineering group.

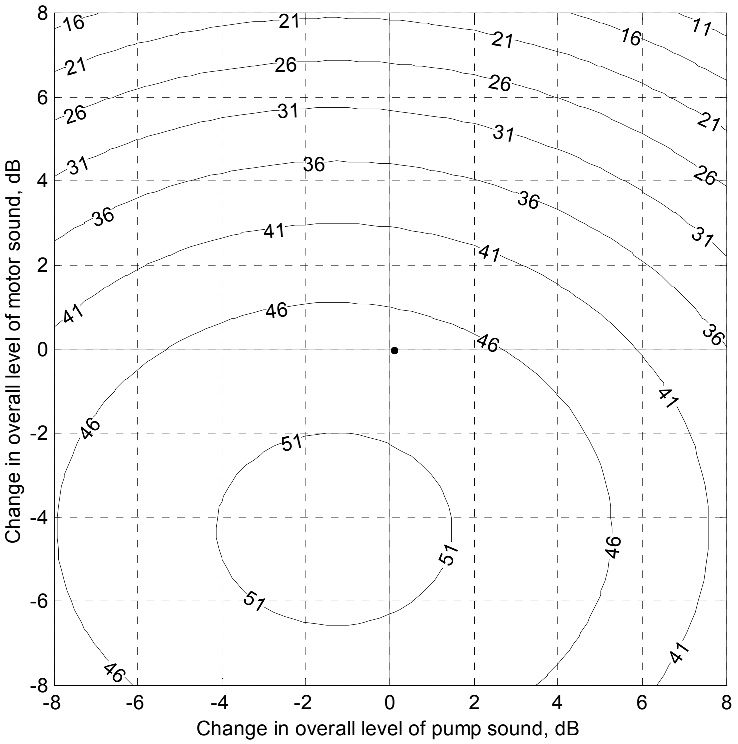

Fig. 1: Regression function for wash cycle perceived quality versus sound levels for pump and motor sounds.

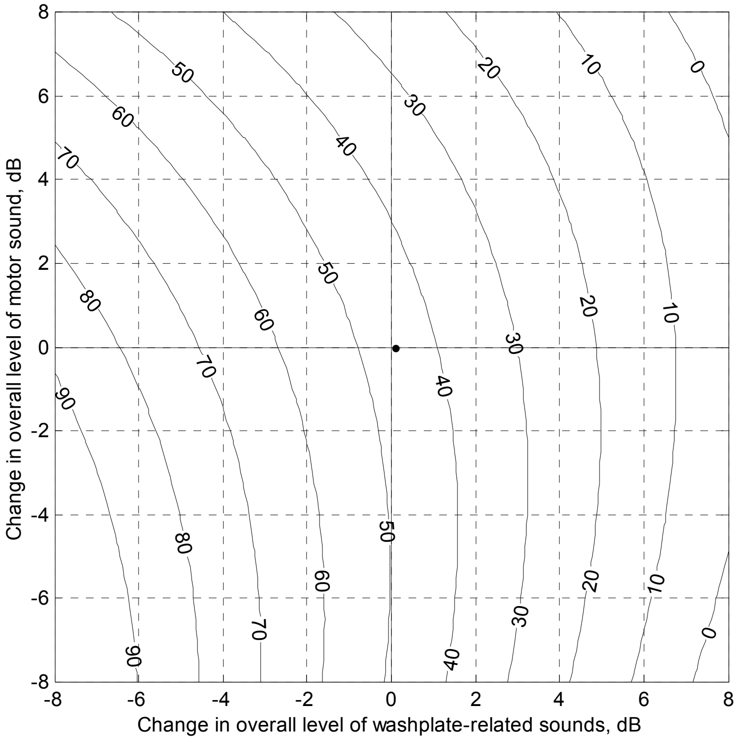

Fig. 2: Regression function for wash cycle "gentleness" rating versus sound levels for washplate and motor sounds.