

Characterization of sound quality of impulsive sounds using loudness based metric

Andrew M. Willemsen and Mohan D. Rao¹

Michigan Technological University, Department of Mechanical Engineering-Engineering Mechanics, 1400 Townsend Drive, Houghton, MI 49931, USA

PACS: 43.66.Lj

ABSTRACT

A study on the characterization of the sound quality of transient sounds via fundamental psychoacoustic measures is described in this paper. Specifically, the overall subjective perception of annoyance for transient sounds was studied. Through magnitude estimation and paired comparison jury evaluation experiments, the subjective annoyance magnitudes of 15 transient sounds were determined. For each sound, several objective psychoacoustic measures were calculated, and using simple linear regression models, the relationships between these objective measures and the subjective annoyance magnitudes were investigated. Examined psychoacoustic measures included loudness, sharpness, roughness, fluctuation strength, tonality, and a new loudness-based measure of impulsiveness. The new impulsiveness measure is based on the summation of the magnitudes of impulse-induced peaks in the loudness time history for a sound (calculated according to DIN 45631/A1). The models were analyzed using several statistical measures of model significance and fit. It was found that for the transient sounds studied, significant relationships existed between subjective annoyance and each of the following psychoacoustic measures: loudness, sharpness, roughness, and loudness-based impulsiveness. These four measures were then combined into a single model for predicting subjective annoyance using multiple linear regression analysis. It was found that this model was highly correlated to the subjective annoyance of transient sounds.

¹ Currently on leave at the Petroleum Institute, Abu Dhabi, UAE

INTRODUCTION

Consumer perception and satisfaction with a particular product is largely dependent on the sound characteristics of the product. Thus, during product development, sound quality is an important design consideration. However, evaluating sound quality in a meaningful and practical manner is often a difficult task. Namely, sound quality is dependent on human perception, and thus must typically be assessed subjectively by a listening panel. Subjective evaluation of sound quality has several drawbacks including:

- Time and costs involved in recruiting and training test subjects and performing the evaluation.
- Implementation only practical at the end of the design process
- Poor repeatability and consistency of results
- Difficulty in extending results to quantifiable design targets
- No insight provided into what individual sound attributes contribute to the overall impression of sound quality

An *objective* measure, which could quantitatively assess the impression of sound quality, could potentially eliminate these drawbacks, making improvement of product sound quality a more realistic design goal. Essentially, such an objective measure would make it possible to replace a human listening panel with a conventional microphone or binaural head.

Many psychoacoustic metrics have been developed to quantify the subjective perception of particular sound characteristics. These measures were developed through extensive subjective evaluations and are meant to simulate the sound processing of the human hearing system. Some common psychoacoustic metrics include loudness, sharpness, roughness, fluctuation strength, and tonality [1]. However, each of these psychoacoustic metrics only quantifies individual sound attributes that combine to give the overall impression of sound quality. There is no widely accepted single psychoacoustic metric which completely characterizes overall sound quality. The objective of the present study is to develop such an *overall* sound quality metric. Overall sound quality can be subjectively described in terms of one of several qualities, commonly including “annoyance”, “pleasantness”, or something more product specific such as “sportiness” or “luxuriousness” for automobiles [2,3]. Studies have shown that the choice of terminology can have a significant effect on the resulting assessment of overall sound quality [3], so the current study specifically examines sound quality in terms of perceived annoyance. A composite psychoacoustic metric of perceived annoyance is, thus, developed; combining separate psychoacoustic measures of different sound attributes into one all encompassing measure.

This study was specifically focused on the perception of sound quality for impulsive sounds. Impulsive sounds are one type of transient sound characterized as having short-duration increases in amplitude, occurring at high rates of change [4]. This type of sound event is often described as “clicks”, “squeaks”, “rattles”, and “pops”. Impulsive sounds are common in information technology (IT) products and devices, and many of the sounds examined in this study are from IT devices [5]. Human hearing is particularly sensitive to transient sound characteristics, such as impulsive content [6]. Thus, a composite measure of annoyance for impulsive sounds must not only account for the standard psychoacoustic quantities (listed above), but also some measure of the degree of impulsive content within a sound. Thus, the first aim of the presented study is to develop a measure of impulsiveness which quantifies the perception of impulsive content within a sound in an accurate, yet simple, manner. The methodology used to first develop the impulsiveness measure and then the perceived annoyance measure for impulsive sounds is described herein, and is an extension of a previous publication [7].

LOUDNESS-BASED IMPULSIVENESS MEASURE

The “impulsiveness” of a sound refers to the degree of impulsive content *perceived* in the sound. It is related to both the perceived magnitude and number of impulses the sound contains. Several standardized [8,9] and non-standardized [10-12] methods have been developed to quantify impulsiveness, but most of these existing measures are either quite complex or have been found to be inadequate. The inadequacy of some of these existing measures often stems from their basis on purely physical measures of a sound, which do not adequately account for human perception of sound. For IT products, a commonly used impulsiveness measure is the impulsive parameter as specified in ISO 7779 [8]. The ISO 7779 impulsive parameter is simply the difference between the a-weighted sound pressure level with and without impulse-time-weighting applied. Not only is sound pressure level insufficient for measuring the perceived magnitudes and durations of impulsive events, but the slow decay of the impulse-time-weighting hinders the detection of closely spaced impulses within a sound recording, as is demonstrated in Figure 1. Of the numerous impulse-related sound pressure peaks in the recording of the keyboard typing sound, the impulse-time-weighted sound pressure level only detects a handful.

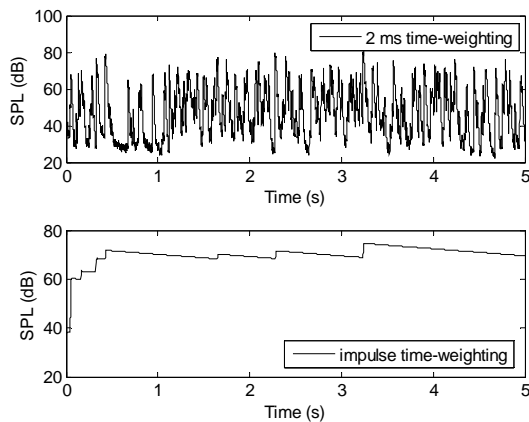


Figure 1. Sound pressure level of impulsive keyboard typing sound for 2 ms time-weighting (top) and impulse time-weighting (bottom)

Thus, before developing a composite measure of annoyance for impulsive sounds, a new measure to quantify the impulsiveness of a sound in an adequate, yet simple manner was first designed. Several criteria for the design of a new impulsiveness measure were established, which are as follows:

1. The measure should only increase in value due to the presence of sudden, short-term transient sound events (impulses).
2. The measure should have a value of zero for completely stationary, non-transient sounds.
3. The measure should not increase in value due to the presence of slow transient sound events.
4. The measure should increase in value for every increase in the number of audible impulses present in the sound.
5. The measure should increase in value as the magnitude of any impulse within the sound increases.
6. The measure should accurately account for the actual perceived magnitude and duration of the impulses.
7. The measure should be independent of the non-impulsive level of the sound.
8. The measure should be independent of the duration of the sound recording.
9. The measure should be independent of the time-resolution of the sound recording.

Additionally, several assumptions were made to simplify the design of the measure, which are as follows:

1. The perception of impulsiveness varies linearly with respect to the magnitude of the impulses within the sound
2. The perception of impulsiveness varies linearly with respect to the number of impulses within the sound
3. The duration of an impulsive sound event is less than or equal to one second
4. The spectral content of an impulsive sound does not affect its perceived impulsiveness

Each of the design criteria and assumptions were satisfied by the new “loudness-based impulsiveness” measure. As its name suggests, the loudness-based impulsiveness uses loudness as a function of time to compute the degree of impulsive content within a sound. Specifically, the “time-varying loudness” as specified by the draft standard DIN 45631/A1 is used [13]. Time-varying loudness adequately represents the human perception of magnitude and duration of sound events. The algorithm specified by the DIN 45631/A1 standard results in a sampled loudness versus time signal for the sound with M data samples (time resolution dependent on the resolution of the original sound pressure signal). The loudness-based impulsiveness measure is then given by

$$I_N = \frac{1}{M} \sum_{i=1}^M [(N_i - N_{b,i})] \quad [1]$$

where N_i is the instantaneous loudness at data sample i of the loudness versus time signal. The term $N_{b,i}$ is the loudness of the non-impulsive components of the sound at data sample i . This non-impulsive, or baseline, loudness is calculated from the 95th percentile of the loudness over a moving 1-second block of time. One second is used since it was assumed that the duration of an impulsive event is less than one second (assumption 3 above). By this assumption, the loudness of a sound must return to its non-impulsive (baseline) level at some point over a designated 1-second block of time, and, thus, the 95th percentile over the 1-second block of time will reflect the baseline value of loudness. The 95th percentile over the 1-second block is used rather than the minimum since the minimum would be affected by instantaneous drops in loudness which may be present in the recording that do not

accurately reflect the true baseline loudness magnitude. This one-second block contains the number of data samples equivalent to one second based on the time resolution of the loudness signal. For each data sample i , the baseline loudness is calculated with the moving 1-second block centered at data sample i . At data samples within 0.5 seconds of the beginning or end of the sound recording, it is not possible to symmetrically center a 1-second block. Instead, the duration of the moving block is reduced from one second to the longest duration block allowable if the block is to remain symmetrically centered at the data sample of interest. Essentially, the baseline loudness will follow the loudness of the non-impulsive content within the sound, and subtracting the baseline loudness signal from the original loudness signal leaves a loudness signal with only impulsive content and a baseline loudness of zero. This is illustrated in Figure 2. The background loudness calculation is given by

$$N_b(t) = \begin{cases} P_{95}[N_0, \dots, N_{2+i_t}], & t < 0.5 \\ P_{95}[N_{t-0.5}, \dots, N_{t+0.5}], & 0.5 \leq t \leq T - 0.5 \\ P_{95}[N_{t-T}, \dots, N_{t_T}], & t > T - 0.5 \end{cases} \quad [2]$$

where i_t is the data sample at time t , T is the total duration of the sound recording, and P_{95} is the 95th percentile of the values enclosed in the brackets.

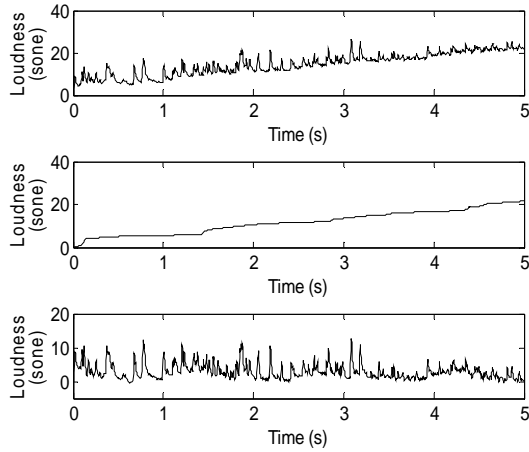


Figure 2. Baseline loudness elimination procedure. Original sound has a slowly increasing transient component and impulsive components (top). Baseline loudness, as calculated by Eqn. 2, follows slowly increasing transient (middle). Subtracting the baseline loudness from the original loudness leaves only the impulsive components (bottom).

This new loudness-based impulsiveness measure was incorporated in the objective annoyance measure developed in this study. As will be discussed later in this paper, the relationship between loudness-based impulsiveness and perceived annoyance for impulsive sounds was found to be much stronger than the relationship between the ISO 7779 impulsive parameter and perceived annoyance.

EXPERIMENTAL METHODOLOGY

In order to develop an objective measure of annoyance for impulsive sounds, correlations between objective psychoacoustic metrics and subjective annoyance ratings for a sample set of impulsive sounds were examined. The impulsive sounds were recorded and prepared, and then an objective psychoacoustic analysis was performed on each sound. Finally, a subjective annoyance evaluation

experiment was conducted to obtain the subjective annoyance ratings for each sound.

Recording and Preparation of Sound Samples

A total of twelve sounds were recorded for the experiment. Since the focus of this research is on the characterization of impulsive sounds, all of these sounds had some degree of impulsive content. Table 1 lists descriptions of the sounds used for the subjective evaluation. Nine of these sounds were various noises emitted by printers during printing, self-maintenance, and self-calibration. Three of the sounds were typing sounds from three different computer keyboards. Together, these twelve sounds were chosen as examples of impulsive sounds common for IT devices.

Table 1. Descriptions of sounds used in the experiment

Sound Number	Description
1	Printer roller loading noise 1
2	Printer roller loading noise 2
3	Printer gear drive calibrating itself
4	Complete printing noise 1
5	Printer solenoid activated latch turning on
6	Computer keyboard typing sound 1
7	Computer keyboard typing sound 2
8	Computer keyboard typing sound 3
9	Complete printing noise 2
10	Complete printing noise 3
11	Printer roller loading noise 3
12	Complete printing noise 4

All sounds were recorded in an anechoic chamber. The nine printer sounds were recorded using a HEAD Acoustics HMS III binaural head and frontend system with an independent-of-direction (ID) recording equalization applied. The recording equalization was used to make the binaural head recordings comparable to conventional microphone recordings, which is necessary when performing objective psychoacoustic analysis. The three keyboard sounds were recorded with a HEAD Acoustic HSU II binaural head and equalizer with a free field (FF) recording equalization applied. The keyboards were placed in front of the binaural head at a normal position for keyboard use, approximately 50 cm from the center of the binaural head. The HSU II binaural head was connected to a 24-bit soundcard (M-Audio, Audiophile USB soundcard) for recording. Both the printer and keyboard sounds were recorded with a 24-bit quantization level and 48,000 Hz sampling frequency.

To prepare the sound samples for the analyses, a representative 5-second segment was selected from each recording. This was done to ensure the results were not affected by differences in duration between the 12 recordings. For the objective analysis, only the channel with the largest overall sound pressure level for each two-channel binaural recording was analyzed. For the subjective annoyance analysis, the sound samples had to be properly equalized so that the sounds the test subjects heard through a set of headphones was equivalent to the sounds they would hear if actually present in the recording room. The playback equalization in part needed to remove the effects of the original recording equalization applied to each sound, so

different equalizations were applied depending on if an ID or FF recording equalization was originally applied.

Objective Psychoacoustic Analysis

A number of common psychoacoustic metrics were computed for each of the twelve sound samples including:

- Loudness – time-varying loudness specified by draft standard DIN 45631/A1 [13]
- Sharpness – Aures method applied to time-varying loudness [14]
- Roughness – “Partial roughness” calculation implemented in HEAD Acoustics Artemis software
- Fluctuation Strength – Hearing Model based calculation implemented in HEAD Acoustics Artemis software
- Tonality – Aures and Terhardt model of tonality
- ISO 7779 impulsive parameter [8]
- Loudness-based impulsiveness

With the exception of the two impulsiveness measures, each psychoacoustic metric was computed as a function of time. This is because all of the sounds had time-varying characteristics, and only examining a single *average* value for these metrics may not have accurately reflected their overall perception. However, to examine the relationships between these metrics and perceived annoyance, a singular value was needed to represent each metric over the entire duration of the sound. Several singular values were computed for each sound including average, median, maximum, minimum, 5th percentile, 10th percentile, and 90th percentile. It was later determined which of these singular values best represent the perception of the temporal characteristics of impulsive sounds and produce the strongest correlation to perceived annoyance.

Design of Subjective Evaluation Experiment

To obtain subjective ratings of the perceived annoyance of the twelve sound samples, a jury evaluation experiment was performed. The recruited test subjects were asked to perform two subjective evaluation tasks. The first of these tasks was an annoyance magnitude estimation experiment. For this experiment, an interactive computer interface was designed which allowed the test subjects to listen to each of the twelve sounds (played through a set of headphones) by clicking a corresponding button, and then rate the level of annoyance for each sound by moving a slider bar along a scale. The interface is shown in Figure 3. The scale was not given any numerical values, but instead has the descriptors “extremely annoying” and “no annoyance” at the extremes of the scale. Test subjects were allowed to listen to the sounds in any order and any number of times. The ratings could be readjusted as many times as the test subject deemed necessary. At the conclusion of the experiment the computer interface output numerical values ranging from 0 (“no annoyance”) to 100 (“extremely annoying”) based on the positions of the slider bars.



Figure 3. Computer interface for annoyance magnitude estimation experiment

The second task performed by the test subjects was a paired comparison experiment. For this experiment, all possible pairs of the twelve sounds were presented, and the test subjects were asked to select one sound for each pair which was the most annoying. Additionally, fifteen of the pairs were repeated, with the order of the sounds within the pair switched. These repeated pairs were used to determine the test subjects' judgment repeatability, as well as to determine if the test subjects' judgments were affected by presentation order. In total, 81 pairs of sounds were presented to the test subject. The pairs were presented in a random order and the ordering of the sounds within the pairs was also random (with the exception of the fifteen repeated pairs). This was done to ensure the results were unaffected by presentation order. Sounds were presented with a three second pause between pairs and a two second pause between sounds within a pair. The sounds in each pair were only played once, but the test subjects were already familiarized with the sounds by the preceding magnitude estimation experiment. An interactive computer interface was implemented to conduct the paired comparison experiment and record the responses of the test subjects.

A total of 36 test subjects (32 male and 4 female) were recruited for the experiment. The eligible test subjects were between 18 and 30 years of age (average age of 22) and had no known hearing impairments or learning disabilities. Additionally, each test subject was required to pass an audiometric screening to ensure their hearing thresholds in the range of 250 Hz to 8000 Hz were normal.

The total duration of the test was approximately 45 to 60 minutes. The test was conducted in a small office with walls treated to be acoustically absorptive. A printer was placed on the table next to where the test subject sat. The office environment and printer were used to simulate the environment where most of these sound samples would normally be heard. Studies have shown that the accuracy of subjective sound quality evaluations increases when the listener is located in the actual sound environment [15]. The test environment is shown in Figure 4.



Figure 4. Testing room for conducting the jury evaluation experiment

RESULTS AND DISCUSSION

Results of Subjective Evaluation

The results of the subjective annoyance evaluation experiment were used to obtain a set of annoyance ratings for the twelve impulsive sound samples. However, before deriving the annoyance ratings from the experimental results, the performance of each test subject was analyzed, and the results of any poorly performing test subject were removed from the experiment. In particular, the repeatability and consistency of each test subject's judgments were determined. Both of these performance measures were based on the results of the paired comparison test. Consistency in a paired comparison test refers to how well the test subject's individual pair judgments make sense when examined together. For example, if the pair judgment between sound A and sound B yields that sound A is more annoying than sound B ($A > B$), and the pair judgment between sound B and sound C yields that sound B is more annoying than sound C ($B > C$), then it would be expected that in judging between sounds A and C that sound A would be more annoying than sound C ($A > C$). However, if the test subject ranks $C > A$, this is known as a circular triad and is inconsistent. The Kendall consistency of each test subject was calculated from

$$\zeta = \left(1 - \frac{24c}{t(t^2 - 1)}\right) * 100\% \quad [3]$$

where ζ is the consistency, c is the number of circular triads in the results of the test subject, and t is the number of sounds in the experiment [16]. The repeatability of the test subjects' judgments was determined from the percentage of the fifteen repeated pairs judged the same during both presentations. Only test subjects having a consistency greater than 70% and repeatability greater than 60% were included in the study [15]. A total of 26 test subjects met these criteria (average consistency of 86.5% and average repeatability of 79.2%).

The annoyance ratings computed based on the results of the remaining 26 test subjects are plotted in Figure 5. For the magnitude estimation experiment, the median values of the magnitude estimates on the 0 to 100 scale are presented along with the interquartile range of the estimates for each sound. The median was used rather than the mean since the median is less affected by outlying evaluations. For the paired comparison experiment, the Bradley-Terry model was used to derive linearly scaled merit values based on the paired probabilities for all sound pairs [16]. The paired probability

is the probability that a listener would find a particular sound of a specified set of two sounds to be more annoying. If the paired probabilities for the two sounds in a pair are near 0.5, their merit values (or scaled annoyance values) are similar. If the paired probabilities are near zero and one, the merit values of the two sounds will be far apart on the merit scale. The merit scale resulting from the Bradley-Terry model estimation ranged from -9.18 to 0, so the merits were re-scaled to fit the same range as the magnitude estimation results (0 to 100). It was observed that the resulting annoyance ratings from the magnitude estimation experiment and paired comparison experiment were relatively similar. All the paired comparison ratings were well within the interquartile ranges of the magnitude estimation ratings. It was thus concluded that the experimental procedures did not have a significant effect on the resulting annoyance ratings, and the experimental design was valid. For the remainder of the study, only the annoyance ratings resulting from the paired comparison experiment were used.

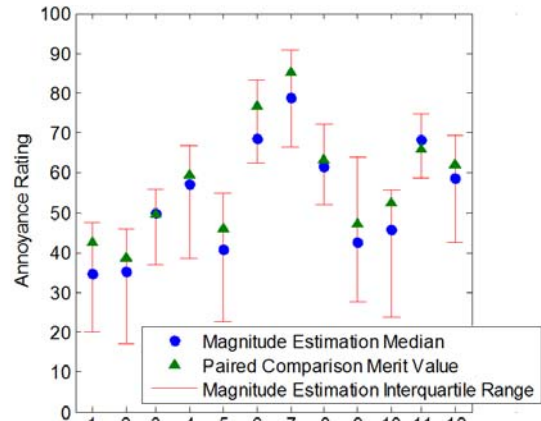


Figure 5. Subjective annoyance ratings from magnitude estimation experiment and paired comparison experiment

Simple Linear Relationships between Sound Attributes and Annoyance

Before developing the composite measure for perceived annoyance, the individual relationships between each of the calculated psychoacoustic metrics and the subjective annoyance ratings were examined. These relationships were derived from simple linear regression models, using the psychoacoustic metric singular values as predictor variables. The strength and significance of each regression model were then analyzed to determine which metrics correlate best to perceived annoyance [17]. The strength of each model was determined based on:

- Coefficient of determination (R^2) – How well the model fits the perceived annoyance ratings. Values near one indicate a strong model.
- Prediction sum of squares ($PRESS$) – How well the model predicts the perceived annoyance ratings. Smaller values indicate a stronger model.
- Spearman's rank correlation coefficient (ρ) – How well the model predicts the perceived annoyance rankings. Values near 1 indicate a strong model.

The significance of the regression relationships were determined based on the 95% confidence intervals for the slope of each model. If the confidence interval of the slope included zero, it was concluded that the linear regression relationship was insignificant.

Table 2 presents the estimated slope and intercept of the fit models along with the values of the model strength measures. The coefficient b_0 is the intercept of the model, and the coefficient b_1 is the slope of the model. These are least squares estimates of the true intercept and slope, β_0 and β_1 , respectively. Table 3 presents the confidence intervals for the slope estimates of the same models. Only the best correlated singular value for each psychoacoustic metric is reported, which are as follows:

- Loudness – 5th percentile, N_5
- Sharpness – median, S_{50}
- Roughness – average, R_{ave}
- Fluctuation Strength – average, F_{ave}
- Tonality – average, tu_{ave}
- Impulsiveness – loudness-based impulsiveness, I_N (ISO7779 impulsive parameter also reported for comparison)

Table 2. Estimated model parameters and measures of model strength for each psychoacoustic metric

Predictor Variable	b_0	b_1	R^2	PRESS	ρ
N_5	23.9	3	0.56	1452	0.79
S_{50}	16.2	24	0.64	1073	0.69
R_{ave}	40.7	6.2	0.81	532	0.83
F_{ave}	48.3	47.1	0.25	2135	0.49
tu_{ave}	60.2	-20.8	0.03	3433	0.20
I_N	33.5	10.4	0.79	655	0.78
ISO7779	53	0.5	0.01	2896	0.17

Table 3. Confidence intervals for the slopes of the models for each psychoacoustic metric (intervals highlighted in gray include zero)

Predictor Variable	b_1	β_1 95% Confidence Limit
N_5	3	[1.1, 4.8]
S_{50}	24	[11.4, 36.6]
R_{ave}	6.2	[4, 8.3]
F_{ave}	47.1	[-9.8, 104.1]
tu_{ave}	-20.8	[-112.8, 71.3]
I_N	10.4	[6.7, 14.2]

Based on the model strength measures, roughness and loudness-based impulsiveness appear to be strongly correlated to perceived annoyance for impulsive sounds. It was also observed that loudness-based impulsiveness is much more strongly correlated to perceived annoyance than the ISO 7779 impulsive parameter, which was found to have no significant relationship to annoyance. Loudness and sharpness were also found to be correlated to perceived annoyance, but to a lesser degree. The measures of fluctuation strength and tonality do not appear to have any significant correlation to perceived annoyance. This is further confirmed by the confidence intervals of the slopes of these two models. The confidence intervals of the slopes of the fluctuation strength and tonality models both include zero. Based on these results, only 5th percentile loudness, median sharpness, average roughness, and loudness-based impulsiveness were considered for inclusion in the composite objective measure of perceived annoyance for impulsive sounds.

Development of Objective Annoyance Measure

Finally, a composite measure of annoyance for impulsive sounds was developed by combining the set or a particular subset of the four psychoacoustic metrics found to be significantly correlated to perceived annoyance on an individual level. The psychoacoustic metrics were combined via multiple linear regression techniques. Each metric was considered a potential predictor variable in a model of perceived annoyance. In addition to the four individual metrics, the interaction of loudness and sharpness was also considered as a potential predictor variable. This interaction variable is simply the product of 5th percentile loudness and median sharpness. This interaction variable was considered since it has been shown that sharpness is not independent of loudness [14]. All possible subsets of these predictor variables were used to model perceived annoyance, and the strength of each model was determined. The same measures of model strength that were previously used to analyze the single linear regression models were again used to analyze the multiple linear regression models. Additionally, two other strength measures were utilized which penalize for using an excessive amount of predictor variables. These were adjusted R^2 , which is a measure of model fit, and Mallows's C_p criterion, which measures model bias [17]. For R^2_{adj} , values near one indicate a strong model, and for C_p , smaller values indicate an unbiased and simpler model.

Table 4. Measures of model strength for various subsets of predictor variables (top three values for each strength measure and the best subset of variables are highlighted in gray)

Predictor Variables	R^2	R^2_{adj}	ρ	PRESS	C_p
$N_5, S_{50}, R_{ave}, I_N, N_5 * S_{50}$	0.939	0.888	0.951	953	6.0
$N_5, R_{ave}, I_N, N_5 * S_{50}$	0.923	0.879	0.972	819	5.5
$N_5, S_{50}, R_{ave}, N_5 * S_{50}$	0.933	0.895	0.979	427	4.5
$N_5, S_{50}, I_N, N_5 * S_{50}$	0.935	0.897	0.965	419	4.4
$N_5, S_{50}, R_{ave}, I_N$	0.935	0.898	0.972	748	4.4
$N_5, I_N, N_5 * S_{50}$	0.923	0.894	0.972	363	3.5
$R_{ave}, I_N, N_5 * S_{50}$	0.923	0.895	0.972	317	3.5
$S_{50}, I_N, N_5 * S_{50}$	0.927	0.899	0.965	330	3.2
N_5, S_{50}, I_N	0.934	0.910	0.965	297	2.4
N_5, R_{ave}	0.911	0.891	0.916	280	2.8
$I_N, N_5 * S_{50}$	0.923	0.906	0.972	273	1.5

Table 4 shows the strength measures for several of the examined multiple linear regression models. Only the models which had one of the best three values (highlighted in gray) for at least one of the strength measures are shown. It was observed that the model which included loudness-based impulsiveness and the interaction of loudness and sharpness had one of the best values for all but one of these strength measures. However, this model did not include roughness as a predictor variable, which was found to be the psychoacoustic metric most correlated to annoyance on an individual level, as indicated previously in Table 2. This is because average roughness was highly correlated to the other included predictor variables, as shown by the correlation coefficients in Table 5; thus, its inclusion would not significantly improve the model. However, the correlations between the predictor variables is reduced if 5th percentile roughness is used in place of average roughness, as shown in Table 5. The correlation of 5th percentile roughness to perceived annoyance on an individual level is nearly as

strong as for average roughness. Thus, another model was fit using 5th percentile roughness (R_5) along with the loudness-based impulsiveness and loudness-sharpness interaction variables. This model's fit to ($R^2 = 0.949$ and $R^2_{adj} = 0.930$) and prediction of ($PRESS = 201$ and $\rho = 0.986$) the subjective annoyance data was significantly improved over the model without roughness. This model was chosen as the objective measure of perceived annoyance for impulsive sounds. The model is given by

$$\hat{A} = 27.73 + 1.24 * I_N + 0.86 * N_5 * S_{50} + 1.81 * R_5 \quad [4]$$

Table 5. Correlation coefficients between the psychoacoustic predictor variables in annoyance model

	$N_5 * S_{50}$	I_N	R_{ave}	R_5
$N_5 * S_{50}$		0.79	0.86	0.72
I_N	0.79		0.91	0.89
R_{ave}	0.86	0.91		
R_5	0.72	0.89		

Figure 6 plots the annoyance ratings predicted by Equation 6 and observed from the subjective evaluation experiment for the twelve test sounds. All the predicted annoyance ratings are relatively close in value to the observed ratings. The observed ratings for all but two of the sounds fall within the 95% confidence intervals for the predictions.

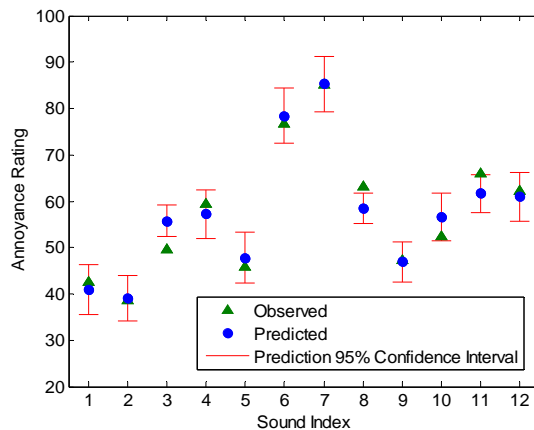


Figure 6. Perceived annoyance predicted by model with loudness-based impulsiveness, loudness-sharpness interaction, and 5th percentile roughness as predictor variables. Observed annoyance ratings and 95% confidence intervals for predictions are also shown.

CONCLUSIONS

Incorporating improved sound quality into product design is essential for improving user perception and satisfaction with a product. However, subjectively evaluating sound quality can be cumbersome and costly, and is generally not well-suited for incorporation in the design process. The objective measure of perceived annoyance developed in this study is a more practical solution for sound quality design. This objective measure, designed specifically for sounds with impulsive content, was found to accurately model the annoyance ratings experimentally measured by subjective evaluation. The annoyance measure is a composite of several fundamental psychoacoustic quantities including loudness (5th percentile), sharpness (median), roughness (5th percentile), and impulsiveness. Since existing measures of impulsiveness were found to inadequately account for the

human perception of impulsive sound content, a new loudness-based impulsiveness measure was developed for inclusion in the annoyance composite measure. The loudness-based impulsiveness measure was designed to more accurately account for the perceived magnitude and duration of impulsive sound events, and yet remain simple to compute. It was found that loudness-based impulsiveness had a much stronger relationship to the perceived annoyance of impulsive sounds than the current standardized measure of impulsiveness for IT devices (ISO 7779). It is recommended that the derived objective measure of perceived annoyance for impulsive sounds, in conjunction with loudness-based impulsiveness, be used for assessing and improving the sound quality of products which emit sounds with significant impulsive content, including many IT devices.

ACKNOWLEDGEMENTS

Financial support for this study was provided by Xerox Corporation. The authors would like to especially thank Jasper Wong of Xerox for his technical support and guidance throughout this study.

REFERENCES

1. H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models* (Springer-Verlag, Berlin, 2007)
2. K. Noumura and J. Yoshida, "Perception Modeling and Quantification of Sound Quality in Cabin" *Proc. SAE 2003 Noise and Vibration Conference* Traverse City (2003)
3. F. Rossi and A. Nicolini, "Squeaking noise psychoacoustic evaluation for car passengers" *Proc. International Congress on Sound and Vibration 15 Daejeon* (2008)
4. T. H. Pedersen, "Objective method for measuring the prominence of impulsive sounds and for adjustment of L_{Aeq} " *Proc. Inter-noise 2001 The Hague* (2001)
5. T. Baird, N. Otto, and W. Bray, "Impulsive Noise of Printers: Measurement Metrics and Their Subjective Correlation" *Proc. Noise-con 2005 Minneapolis* (2005)
6. R. Sottek and K. Genuit, "Models of signal processing in human hearing" *Int. J. Electron. Commun.* **59**, 157-165 (2005)
7. A. M. Willemsen and M. D. Rao, "Prediction of subjective annoyance for transient sounds using a novel loudness-based impulsiveness measure" *Proc. Noise-con 2010 Baltimore* (2010)
8. International Organization for Standardization, *Acoustics – Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment* (ISO 7779:1999, Geneva, 1999)
9. NORDTEST, *Acoustics: Prominence of Impulsive Sounds for Adjustment of L_{Aeq}* (NORDTEST Method NT ACOU 112, 2002)
10. W. Bray, "Crest factor and short-duration transients: influence of environmental background, event duration and measurement time weightings" *Proc. Inter-noise 2006 Honolulu* (2006)
11. M. Blommer, A. Eden, and S. Amman, "Sound Quality Metric Development and Application for Impulsive Engine Noise" *Proc. SAE 2005 Noise and Vibration Conference* Traverse City (2005)
12. R. Sottek, P. Vranken, and G. Busch, "Ein Modell zur Berechnung der Impulshaltigkeit" *Proc. DAGA 1995 Oldenburg* (1995)
13. Deutsche Institut für Normung e. V., *Berechnung des Lautstärkepegels und der Lautheit aus dem*

Geräuschspektrum – Verfahren nach E. Zwicker – Änderung 1 (DIN 45631/A1, 2007)

14. W. Aures, “Berechnungsverfahren für sensorischen Wohlklang beliebiger Schallsignale” *Acustica* **59**, 130-141 (1985)
15. N. Otto, S. Amman, C. Eaton, and S. Lake, “Guidelines for Jury Evaluations of Automobile Sounds” *Sound and Vibration* **36**, 24-47 (2001)
16. H. A. David, *The Method of Paired Comparisons* (Oxford University Press, New York, 1988)
17. M. H. Kutner, C. J. Nachtsheim, and J. Neter, *Applied Linear Regression Models* (McGraw-Hill/Irwin, New York, 2004)