Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Applications of psychoacoustics to information technology products

Willem Beltman[a], Rina Doherty[b], Eric Salskov[c], Philip Corriveau[c], Doug Gabel[a] and Eric Baugh[d]

[a]Intel Corporation, 2111 NE 25th Avenue, M/S JF2-86, Hillsboro, OR 97124, USA
[b]Intel Corporation, 5200 NE Elam Young Parkway, Hillsboro, OR 97124, USA
[c]University of Twente, Dept. Mechanical Engineering, P.O. Box 217, 7500 AE Enschede, Netherlands
[d]Intel Corporation, 15400 NW Greenbrier Parkway, Beaverton, OR 97006, USA
willem.m.beltman@intel.com

Emerging usage models for computing devices require low acoustic noise, for example in home entertainment systems. Studies have shown that in addition to the overall level, the psychoacoustic aspects must be considered. This paper provides an overview of testing techniques that are used in the information technology industry and outlines two specific case studies. First, an extensive subjective psychoacoustic study was designed and conducted in multiple geographies to determine the annoyance aspects of sound from information technology products when used in a home type environment. Over 200 participants in four countries participated in this carefully controlled experiment and rated typical steady state sounds on a 5 point annoyance scale. The relevant sound quality metrics were extracted and geographical variations quantified. Second, in a paired comparison study the influence of modulation on annoyance was investigated by superimposing different frequency and amplitude modulated sounds onto a baseline. The results indicate that modulation can have a significant effect on end user subjective perception.

1 Introduction

The requirement for higher performance and associated increased cooling requirements makes it more difficult to manage system acoustics. Performance systems typically use one or more fans as a part of their thermal solution. These fans can increase the overall noise of the system so acoustic design needs to be integrated with the thermal system design. Acoustic management is increasingly important, and many system manufacturers are setting more aggressive acoustic targets for quiet desktop and client systems. There are several acoustic standards that outline procedures for acoustic testing. Acoustic targets are driven by performance, cost, size, product positioning and voluntary eco-labeling procedures. Typically, acoustic targets are formulated in terms of sound power. This is a useful metric that allows comparison between systems and allows a good first order measure of end-user acceptability. For notebook PCs, the operator position sound pressure level is also commonly used. However, especially with lower noise levels, there are additional requirements that need to be included to reflect customer preference. The goal of this paper is to provide an overview of methodologies that are used in the IT industry to assess product sound quality. First, the metrics are outlined, then the measurement procedures and subjective test methods are described, and finally the work is illustrated using two case study examples.

2 Metrics

A variety of acoustic metrics exist in the industry. Historically, sound pressure measurements at the operator position were used and reported in dBA (decibels, A-weighted). Recently, sound power measurements have become the primary metric. Eco-labels also use sound power as the central metric that is expressed in BA (bels, A-weighted). However, no single metric captures the entire user experience.

2.1 Sound power

Acoustic testing procedures are well documented and standardized. ISO 7779 [1] and ECMA-74 [2] for instance outline procedures to determine the acoustic emissions of computer and business equipment. ISO 9296 [3] and ECMA-109 [4] describe the procedures to determine the declared noise emission values. The primary quantity to be measured and declared in acoustics of computer and business equipment is the declared emitted sound power level, L_{wAd}, measured per ISO 7779 and reported per ISO 9296. This quantity allows comparing different systems or components and to quantify the overall emitted acoustic energy. The main limitation of the sound power metric is the fact that it is a spatially averaged quantity. It does not give information regarding specific positions.

2.2 Sound pressure

The sound pressure level, measured in dBA, is a supplemental metric to identify the sound pressure levels at the operator and bystander positions. Measurement positions and procedures are outlined in ISO 7779 and ECMA-74. The system or sub-assembly is placed on or next to a table, and the sound pressure levels at the operator and bystander positions are determined. The main advantage of sound pressure experiments is that they can identify acoustic levels at the user position. For example, fans are typically located at the back of systems for this reason. Because of the strong spatial dependency, this metric is not recommended for the comparison of products. In addition, it may introduce confusion when measurement distances are not specified or not used consistently.

2.3 Loudness

The loudness level, in sones expresses the perceived loudness of a signal. Several ways exist to calculate the loudness of a signal. In this paper the ISO 532 procedure is followed [5]. Note that this binaural measurement procedure is not standardized in either ISO 7779 or ECMA-74.

2.4 Tonality

The tonality is a measure of the proportion of tonal components in the spectrum of a signal. In this paper, tonality is expressed in the unit tu and is calculated according to the procedures described by Aures [6] and Terhardt [7]. As an alternative to tonality, Annex D of the ECMA-74 standard describes a procedure for determining whether or not noise emissions contain prominent discrete tones: the prominence ratio method [2]. A discrete tone which occurs together with broadband noise is partially masked by that part of the noise contained in a relatively narrow frequency band, called the critical band, which is centered at the frequency of the tone. Noise at frequencies outside the critical band does not contribute significantly to the masking effect. The width of a critical band is a function of frequency. In general, a tone is just audible in the presence of noise when the sound pressure level of the tone is about 4 dB below the sound pressure level of the masking noise contained in the critical band centered around the tone (threshold of audibility). When using the prominence ratio method, a tone is classified as prominent if the difference between the level of the critical band centered on the tone and the average level of the adjacent critical bands is equal to or greater than 9 dB for tone frequencies of 1 kHz and higher, and by a greater amount for tones at lower frequencies.

2.5 Sharpness

The parameter sharpness has been introduced as a measure for the “sharp” or “shrill” character of a sound. In this paper the calculation is performed according to the Aures procedure, with an ISO 532 method for a free field [8].

2.6 Roughness

Roughness was calculated according to the procedure by Aures [9], to capture the sensation of envelope fluctuations at rates between 20 and 300 Hz. The procedure involves calculating partial roughness in sub bands and summing these to obtain the total roughness value.

3 Measurement procedures

3.1 Binaural recording

Binaural recordings are made to collect sound samples for realistic playback. The recording system is typically located in the seated operator position according to ISO 7779, see Figure 1. A special alignment fixture is used to increase the repeatability of the recording. During the recording phase the alignment fixture is removed. All sounds were sampled at 44.1 kHz for durations of 30 seconds.

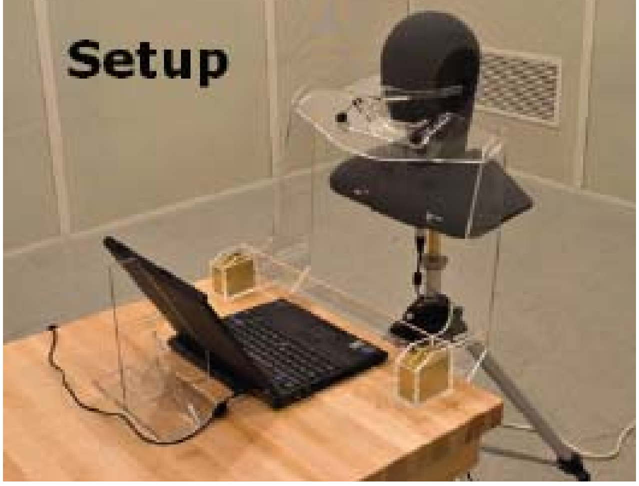

Fig.1 Binaural recording setup with alignment fixture.

3.2 Sample characterization

After the binaural recording, the sounds are characterized in terms of the basic psychoacoustic parameters: sound pressure level, loudness, tonality, prominence ratio, tone to noise ratio, sharpness and roughness using analysis software. This allows to select a sample set in the subjective test procedure.

4 Subjective test procedures

4.1 Test method

Two types of experiments are typically used in subjective assessments: a fixed rating scale experiment or a paired comparison procedure. A fixed rating scale is advantageous for large sample sets. However, it also has the drawback that it does not allow a direct comparison between sound samples. The paired comparison procedure is especially suited for smaller sample sizes, as the amount of tests for comparing all samples to each other quickly grows with sample size. Although this procedure uses a direct comparison between samples, it is possible to extract model parameters using for example the Bradley-Terry probability model using the matrix with relative preference frequency scores.

4.2 Superposition on background noise

The same sample sounds were superimposed on three different background noises. The background noise levels used in the current study represent three different environments of interest and are used to represent typical environments [10]. In addition to the sound samples, the background sound itself was also played as a sample to quantify if participants would appropriately use the rating scale. Background noise levels are of a broadband nature and therefore representative samples with an average broadband contribution were collected. Distinct events, such as a telephone ringing or audible conversation in the office background or a car driving past in the outdoor background, were avoided.

4.3 Test environment

The environment in which the tests are conducted is of critical importance, mainly because of possible noise contamination. The subjective evaluation should therefore be conducted in a quiet jury test environment. The absolute A-weighted sound pressure level in the test environment in the current test cases did not exceed 25 dBA, and preferably should be less than 20 dBA. In addition to the overall level, the sound spectrum was measured in order to verify that no distracting spectral noises, such as pure tones, were present in the test environment. An example of the subjective test setup and environment is given in Figure 2.

.Fig.2 Example subjective test environment.

4.4 Target population selection

In order to get sufficient sampling for statistical analysis, a minimum number of 25 to 30 participants are needed. A larger sample size will positively affect the statistical confidence and enhance the data analysis results. Participants should be selected carefully in order to get a representative sample of the population of interest. Typically, a distinction is made between experienced listeners, for example people inside the company who continually evaluate product sound quality, and inexperienced listeners. In the current case studies, a respected external market research company was hired to recruit a random selection of participants. Special screeners were developed for this purpose.

4.5 Participant instructions

It is important to provide the participants with simple and clear instructions. The instructions can influence the focus of the participants and therefore close attention needs to be paid to giving clear and consistent directions and explanations. After participants are asked to read and sign a consent form they were briefed on the test apparatus and the interface they would use to select their ratings for the sound samples. A practice session was incorporated into the design so as to allow the participants to familiarize themselves with the range of sounds as well as the interface and rating scales. Participants were given ample opportunity to ask questions and were reminded that at any time they needed help to ask the test facilitator immediately.

4.6 Hearing screening

Before the subjective test, an audiogram was measured for each participant to ensure the participants have hearing within an acceptable range. An automated test program was developed for this purpose. The data from the audiograms was recorded, but not shown to the participants. The results from participants with non-normal hearing was excluded from the data analysis.

4.7 Sample repeat

Within every test section, at least one sample was played multiple times. The multiple presentations of the same sample were used to check juror consistency in the rating process.

4.8 Randomized presentation orders

The presentation order of the samples was randomized in order to prevent influence of the presentation order on the subjective ratings. The randomization is done for each session of six to eight jurors; within a session all jurors have the same presentation order as they are all listening to the same playback system. The exact presentation order for each participant was stored in order to allow a detailed statistical analysis of presentation order effects.

4.9 Data analysis

Detailed statistical analysis was performed on the data sets to extract for example quantitative metrics, correlations and confidence intervals.

5 Case study I: Multiple geographies

The objective of this study was to quantify the annoyance to steady state sounds from IT equipment in multiple geographies. For this purpose, tests were conducted in Germany, Sweden, the USA and China [11].

5.1 Test procedure

Because of the large sample set, a fixed 5 point annoyance rating scale was used, see Table 1. Tests were carried out in three listening sessions for each panel, each approximately 15 minutes long. Each session contained the same samples, but in different randomized orders and with the different background levels of 31, 33 and 37 dBA. For this type of experiment it is extremely important that the experimental design is done correctly and hence pilot runs were carried out with both trained and untrained listeners. Data from a total of 207 participants was analyzed, comprising 103 females and 104 males.

5.2 Sample selection

A database with representative sounds was used to generate a sample set. The sounds in the database were all characterized in terms of the basic psycho-acoustic parameters: sound level, loudness, tonality, sharpness and roughness. In the selection of a reduced sound samples set, a minimum number of independent variables should be used. For example, the loudness and the sound level are highly correlated. After careful analysis of the sample database, a reduced set of 66 samples was selected.

Rating	Interpretation
5	Extremely annoying
4	Very annoying
3	Slightly annoying
2	Perceptible but not annoying
1	Not perceptible

Table 1 Subjective rating scale.

5.3 Results

A comparison between the ratings in Germany, Sweden, the USA and China is shown in Figure 3. The results show that, especially for lower noise levels, the variations are significant. This is confirmed by the values for the annoyance scale ratings in the different countries. The background noise levels for each country were used in the subjective ratings, e.g. a background noise level of 31 dBA for Germany, a background noise level of 33 dBA for Sweden and a background noise level of 37 dBA for China and the USA.

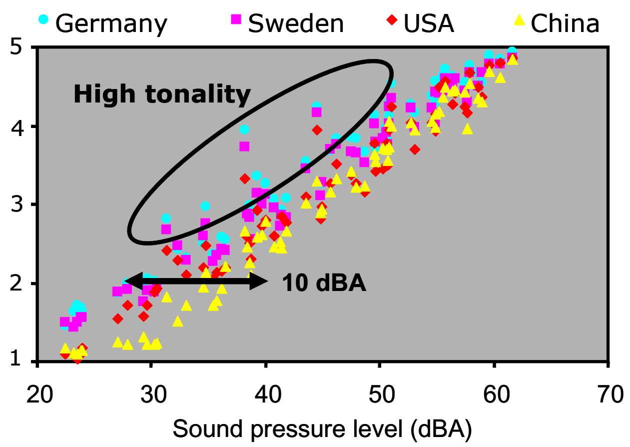

Fig.3 Results of tests in multiple geographies.

It was demonstrated that the main parameters governing annoyance are the sound power level and tonality. As expected, annoyance increases with the sound level. A linear relationship exists between the sound power level and the annoyance on this scale. By comparing the results from different geographies, it was demonstrated that large significant variations exist. For a given sound power level, the highest annoyance ratings were found in Germany and Sweden, the lowest annoyance values were found in China. It was also demonstrated that people can perceive sounds with overall levels below the background noise level.

6 Case study II: Modulation

The objective of this study was to quantify the effect of modulated tone on the acoustic perception. Previous studies have indicated that modulated tones are important in assessing the product sound quality.

6.1 Test procedure

A paired comparison test procedure was used for this experiment. Participants were asked to indicate a preference for sound A or sound B. A tie was not allowed, i.e. a forced choice procedure was adopted. One participant completed the test at a time, *i.e.* the test was conducted serially. An automated test program was used for the hearing test and the paired comparison tests. The hearing threshold curve for each participant was measured and stored. Then, warm up sounds were played to familiarize participants with the interface and voting system. After that the actual test started and the scores were stored in a spreadsheet. The complete test took about 30 minutes per participant. The data for each participant was then read in and used to compile the combined statistics. A data screening was performed to identify data sets that were omitted from the statistical analysis. The statistical analysis was also automated and included operations such as the Bradley-Terry model and Friedmann Dunn statistics.

6.2 Sample selection

A baseline sound was played, on which a tone signal was superimposed. The baseline sound was chosen from the available database of sounds. The sample of interest was of a broadband nature and did not possess tonal or high sharpness characteristics. In the previous study this sample was used, and in that study it received an annoyance rating of 1.89. On a 1 to 5 rating scale this score translates to a value just below “perceptible but not annoying”. Next, modulated sounds were generated to superimpose on the baseline sound. A tone with a carrier frequency of 1 kHz was superimposed on the baseline sound, with varying amplitudes and degrees of amplitude or frequency modulation, see Table 2.

Sample	Level (dBA)	Type	fc (Hz)	fmod (Hz)	m	fd (Hz)
1	30	S	1000	---	---	30
2	35	S	1000	---	---	35
3	40	S	1000	---	---	40
4	35	AM	1000	1	0.3	35
5	35	AM	1000	1	0.5	35
6	30	FM	1000	1	---	30
7	35	FM	1000	1	---	35
8	35	FM	1000	1	---	35

Table 2 Added tone characteristics for modulation study. S=steady state, AM=amplitude modulated, FM=frequency modulated, fc=center frequency, fmod=modulation frequency, m=modulation strength, fd=frequency deviation

6.3 Results

A total of 25 people participated in the test. The results were reviewed to test for individual participant performance and overall stability in the experimental design. Average participant repeatability was high at 83.3%, well above the

60% criteria, indicating that the sample sounds were somewhat easily discriminable. All participants had individual repeatability scores above 60%, except for one participant. Results of the Kendall consistency test revealed that seven out of the 25 participants received a value at or below 75%. The Chi square analysis of the consistency showed that the participants' results were not significantly consistent while all the other participants' preferences were statistically consistent at a significance level of 0.05. Therefore, the participants' data were removed from the data set and not included in further analysis. The data from an additional 2 participants was excluded because of hearing loss. The average Kendall consistency measure for the remaining 16 participants was 91.6%. The Kendall concordance among participants was calculated by using the ranking orders of all the participants. The results confirmed that the participants were in agreement with each other, with a Kendall concordance of 0.7383 ($p < .05$). To test to see if statistically significant differences existed among the sound sample preferences, the Friedman analysis of variance was carried out on the mean ranks for each of the nine sounds. The Friedman test statistic results $X^2(7, N = 16) = 82.7$, $p < .05$ confirmed that there was a significant preference difference across sound samples. To further investigate the difference(s) the Dunn paired test was used, which revealed several pairs of significant mean rank preference differences. The mean ranking of the samples is given in Figure 4.

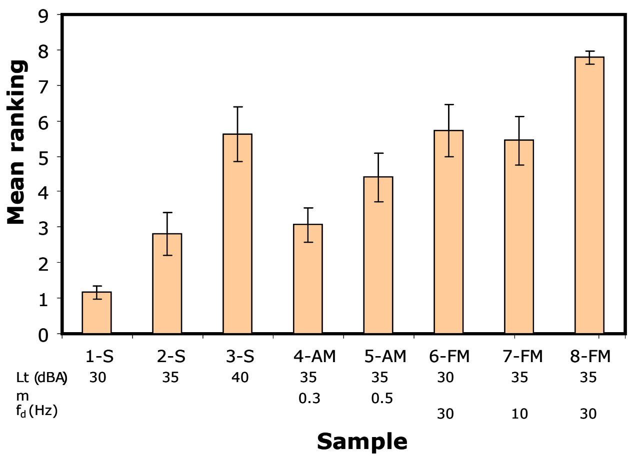

Fig.4 Mean ranking of modulation samples

The results indicated a preference for sound sample 1, as expected. As the tone level is increased (samples 2 and 3), the ranking increases, indicating a lower preference. Samples 4 and 5 are the amplitude modulated sounds. It is interesting that the mean rank for sample 4 is almost the same as that for sample 2. Both samples have the same added tone level of 35 dBA. The modulation of 0.3 for sample 4 apparently does not significantly affect the rating. However, as the modulation is increased to 0.5 the mean rank increases, suggesting a deteriorating sound quality. Unfortunately, because of the relatively large number of screened out data sets, no definitive conclusion can be drawn. From the figure, it is apparent that frequency modulated tones significantly degrade the sound quality. For example, sample 6, with an added tone level of 30 dBA, was rated equal to a stationary tone that was 10 dB higher in amplitude. When the frequency modulation deviation amplitude decreases from 30 to 10 Hz in samples 6 and 7, the ranking decreases slightly even though the level of the added tone increased from 30 to 35 dB. This suggests that the frequency deviation in frequency modulated tones is extremely important.

5 Conclusion

In this paper, an overview of techniques was presented to assess the psycho-acoustic aspects of noise from IT products. The test cases illustrate that geographical differences, the sound level, the tonal content and the presence of modulated tones are key aspects.

Acknowledgments

The authors would like to thank Paul Sorenson, Murli Tirumala, Daryl Nelson, Jessica Gullbrand, Rafael de la Guardia, Jose Cordova and Kaleen Man for their help.

References

[1] ISO 7779:1999 Acoustics -- Measurement of airborne noise emitted by information technology and telecommunications equipment

[2] ECMA-74 Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment, 9th edition (December 2005)

[3] ISO 9296:1988 Acoustics -- Declared noise emission values of computer and business equipment

[4] ECMA-109 Declared Noise Emission Values of Information Technology and Telecommunications Equipment, 4th edition (December 1996)

[5] ISO 532:1975 Acoustics -- Acoustics - Method for calculating loudness level

[6] W. Aures, "A model for calculating the sensory euphony of various sounds", Acustica, Vol. 59(2), 130-141 (1985)

[7] E. Terhardt, "Algorithm for extraction of pitch and pitch salience from complex tonal signals", Journal of the Acoustical Society of America, Vol. 71(3), 679-688 (1982)

[8] W. Aures, "The sensory euphony as a function of auditory sensations", Acustica, Vol. 58(5), 282-289 (1985)

[9] W. Aures, "A procedure for calculating auditory roughness", Acustica, Vol. 58(5), 268-281 (1985)

[10] R. Doherty, E. Salskov, P. Corriveau, P. Sorenson, D. Gabel, W.M. Beltman, "Background Noise Levels in the Home Environment", NoiseCon 2005, Minneapolis

[11] R. Doherty, E. Salskov, P. Corriveau, P. Sorenson, D. Gabel, W.M. Beltman, "Human annoyance levels to PC sounds in the home background noise environment", InterNoise 2006, Honolulu