A NEW PARADIGM IN PSYCHOACOUSTICS, PART FIVE: SOUND QUALITY METRICS

By Steven J. Orfield

In Part Four of this series, the task of sound quality analysis was illustrated via a number of project examples; additionally, some of the politics and corporate dynamics of the sound quality process were discussed.

In Part Five, we will begin to deal with attempts within the psychoacoustic community to analytically quantify the relationship between responses of listeners rating sound quality and the physical measurement of sound itself. For some time, it has been the hope of the psychoacoustic community, and certainly of the product acoustical engineer, that perceptually-based acoustical calculations could be used to solve SQ problems, in lieu of complex listener-based research. Thus, for example, a manufacturer might measure the sound of a lawn mower and then input the spectral and time-based measurements into a computer program to calculate its "Loudness." In lieu of the $50,000-$100,000 often spent to determine the appropriate sound quality standard for a product, this could theoretically be done both less expensively and more quickly on computer, if calculations accurately predicted listener responses to sound samples.

Of specific interest in considering analytical models of sound quality parameters (metrics) is the knowledge, discussed in Part One, of the broad potential base of psychoacoustics. Our recent history in acoustics has seen numerous attempts to calculate and correlate perceived sound quality (positive and negative) results from measurements. Included in this continuum are the fields of audiology and psychology. Included within the acoustic specialties are architectural acoustics, environmental acoustics, product noise, etc.

INCLUDED WITHIN THE ACOUSTIC SPECIALTIES ARE ARCHITECTURAL ACOUSTICS, ENVIRONMENTAL ACOUSTICS, PRODUCT NOISE, ETC.

As has been discussed in Part Two of this series, the SQ listening jury has been assumed to be the basis for validation of sound quality research findings. If the analysis of sound quality measurement suggests one conclusion and the properly designed jury instrument (test) suggests another, then, prima facie, the analysis is considered to be incorrect. (Keep in mind the fact that typical physical acoustic measurements do not correlate well with most jury findings).

There are available, in the cost range of about $100,000+, analysis systems which employ psychoacoustic calculations and SQ "metrics" in order to analytically solve sound quality problems. A listing of the more common metrics is noted below.

Typical Sound Quality Metrics
- *Loudness*
- *Sharpness*
- *Roughness*
- *Fluctuation Strength*
- *Pleasantness*
- *Kurtosis*
- *Tonality*

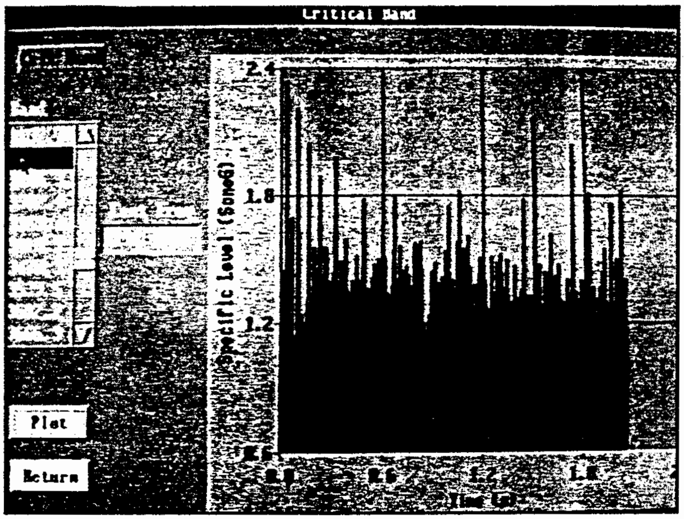
There are also many variations of definitions and calculations for many of the SQ metrics descriptors as the psychoacoustic research underlying the metrics comes from many different sources and finds many points of disagreement. What follows is a look at the Zwicker metrics and calculations. (Throughout this article, I am borrowing terminology and explanations from the Zwicker "Psychoacoustics; Facts and Models" text, as I believe that Zwicker's explanations and descriptions are often the most useful.)

METRIC DEVELOPMENT

The best example of metric development surrounds the extensive efforts of E. Zwicker and H. Fastl to develop psychoacoustic metrics describing human responses to hearing phenomena. Zwicker used audio electronic presentation systems (sound systems) to present stimuli characterizing such

---

Steven J. Orfield is the President of Orfield Associates Inc. in Minneapolis, Minnesota.



Critical band time plot.

descriptions as loudness, sharpness and roughness and fluctuation strength. A summary of much of his work is contained in his last book, "Psychoacoustics; Facts and Models (see Sound & Communication, May 1992 book review).

Generally, the acoustic stimuli presentations underlying the psychoacoustic research of Zwicker were "forced-choice" tests, with the listener hearing two consecutive sounds and rating the relationship between them in terms of the dimension (metric) under consideration. These tests provided an initial basis for the development of relationships between the measured sound and reports of its perceived descriptive qualities.

METRICS WERE DEVELOPED TO DEAL WITH BOTH FREQUENCY-BASED RATING AND TIME-BASED RATING OF SOUND SAMPLES, AS WELL AS TO BE USED IN OVERALL DESCRIPTIONS OF THE LISTENER'S

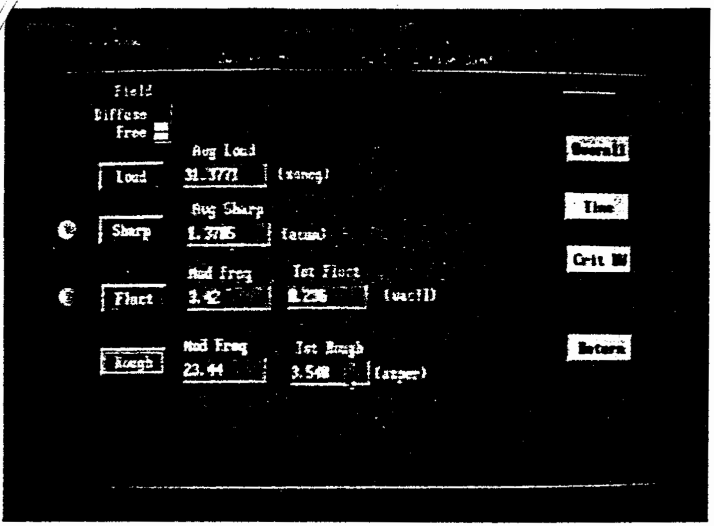
Some of Zwicker's experiments were based on presentations of steady state sounds, and the results of these tests were generally related to average samples of its spectral distribution (average frequency responses). Characteristic of this type of experiment were the use of the Loudness and Sharpness metrics.

Other tests were based on time-varying samples, and the results of these tests were generally related to the rate of change in the sound, in level or in frequency (examples are Roughness and Fluctuation Strength). A third type of metric was developed to summarize complex listener response, and Pleasantness is an example. This metric is calculated using the above set of metrics which describe different parameters of sound quality.

Thus, metrics were developed to deal with both frequency-based rating and time-based rating of sound samples, as well as to be used in overall descriptions of the listener's positive and negative response to a sound sample. Underlying Zwicker's work on SQ metrics, there is a tendency to consider sound in the abstract independent of products. Thus, these metrics do not deal with differences in product sound quality based on issues such as "appropriateness" of the sound sample or other performance-based descriptors, such as "power" of the sound sample.

CONCEPTUAL PROBLEMS WITH METRICS

It is interesting to note that as applications engineers and other researchers began to work with psychoacoustic metrics, they began to intuitively "scale" their responses and to equate level with metric value. They also tend to attempt to correlate their perceived annoyance with level. It is important to note that many metrics, such as sharpness and roughness, may be very low in level and



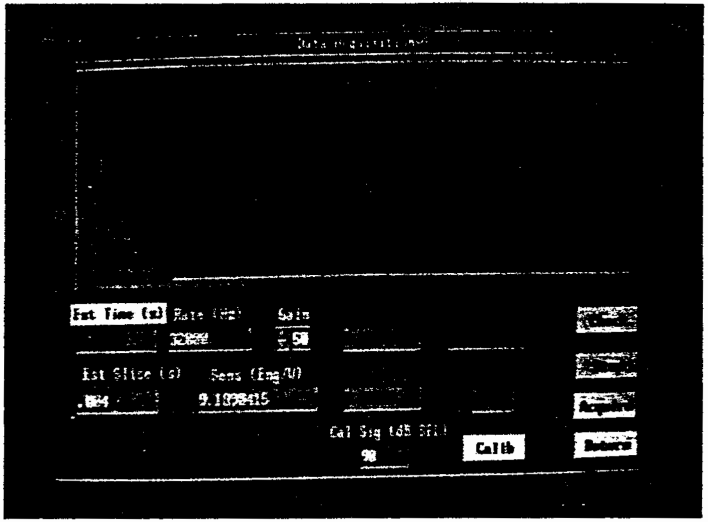
Avg Loud 31.3771 (sones)

Avg Sharp 1.3785 (acum)

Mod Freq 3.42 Int Fluct 0.236 (vacil)

Mod Freq 23.44 Int Rough 3.548 (asper)

Roughness fluctuation.



Live data acquisition.

may be below the threshold of annoyance or negative response. Thus, the experimenter must begin to accommodate the idea that a sound may be sharp but not objectionable.

It is also interesting that there are no commonly available scales with which to evaluate the threshold and limits of listener response to any specific product. This is due partially to the fact that product expectations suggest that one product may be "appropriately" rough (*e.g.*, lawn mower) or inappropriately rough (*e.g.*, food processor). It is due also to the fact that background noise level will change absolute signal-to-noise ratios for the same product under alternative environments. (The demographics, within certain product markets, of hearing loss also have the same resulting changing S/N levels.)

Also, there is a wide variation in user response to product sounds, based on the product performance under unrelated criteria. Finally, marketing efforts can change a buyer's response to product expectations and ratings (*e.g.*, the ongoing sensitization of the population with regard to product noise).

SPECIFIC SOUND QUALITY METRICS

With this general conceptual background in mind, it is instructive to begin to look at the most common individual metrics; these are Loudness, Sharpness, Roughness and Fluctuation Strength. These metrics underlie most of the analytical work currently being undertaken by sound quality engineering staffs and should provide a good introductory look at the application of computer analysis to the sound quality problem. Prior to consideration of these metrics, it is important to mention the concept of critical bands. Critical bands are roughly defined as the band limit of sound adjacent to a given sound which have some potential masking effect of the sound under consideration. Critical bands, in general, describe the bandwidth sensitivity of the ear at various frequencies. Some sound quality calculation, and some new acoustic analysis systems, calculate and measure directly in critical bands rather than octaves or fractional octaves (e.g. $^1/_1$, $^1/_3$, $^1/_6$, $^1/_{12}$ octave).

LOUDNESS

Loudness is generally considered the oldest of the sound quality metrics in use, and it clearly precedes the development of "sound quality" as a formal field. Underlying the loudness metric is long term attempt within the acoustic community to equate psychophysical loudness with the A-Weighted decibel scale. Established in both the legal domain and in acoustic measurement, in general, is the development and use of the A-weighted decibel scale. Based on human hearing sensitivity thresholds in each of a number of frequency ranges (generally, 20 Hz—20 kHz), this scale is used to modify non-weighted (linear) physical acoustic measurements via de- emphasizing those frequencies which have a higher hearing threshold for the typical listener.

Experiments in threshold detection are generally performed via the presentation of only one frequency component (for example, a pure tone at 500 Hz). They do not take into account the perception of broadband sounds (for example a sound covering the 500 Hz octave), nor do they take into account the effect of auditory masking of one sound by another. Thus, two sound samples with a nominally similar A-weighted decibel level may have very different low frequency distributions, and the one with more low frequency content may produce less perception of higher frequency sounds due to "forward" masking.

The dBA scale also does not take into account that findings that the ear's sensitivity to frequency discrimination has been characterized by the development of "critical" bands (*i.e.*, bands of varying bandwidth across the frequency spectrum) describing the average frequency discrimination of the auditory system. With regard to critical bands, specifically, below 500 Hz, the masked threshold is independent of frequency; above this point, the masked threshold increases by 10 dB per decade.

The loudness level of a sound is the sound pressure level of a 1000 Hz tone in a plane wave and presented frontally that is as loud as the sound; the unit of loudness is the "phon." Specific loudness is described in: "sones," with one sone being equivalent to a 1000 Hz 40 dB tone (40 phons). From this general description has been developed a set of equal loudness contours.

This loudness metric has been incorporated into the ISO standards via the use of a $^1/_3$-octave scale and calculation, and in its current embodiment, it does not deal with the time-varying nature of many sounds. Loudness is generally intended to describe the perceived intensity of a time invariant sound sample.

SHARPNESS

Sharpness is a less obvious and more arbitrary descriptive metric. It is defined by the spectral content of narrow band sounds and the spectral envelope, in general. Sharpness is defined for narrow band noises as increasing with increasing frequency. At low frequencies, sharpness increases almost in proportion to the critical band rates. Bandwidth is nominally inversely proportional to sharpness. Sharpness is not very dependent upon level and has little dependance on bandwidth, when considering bandwidths smaller than critical bands.

The strongest parameters determining sharpness are spectral content and center band frequency of narrow band sounds. Narratively, sharpness describes the high frequency nature of the sound sample; the purer the tone and the higher the frequency, the sharper the sound will be rated. The unit of sharpness is the "Acum."

SHARPNESS DEFINITION

A reference sound producing one Acum is a narrow band noise one critical band wide at a center frequency of 1 kHz having a level of 60 dB. The overall spectral envelope is the principal determiner of sharpness. Metrics used in subjective testing which are intended to describe sharpness include Shrill, High Frequency, etc.

ROUGHNESS

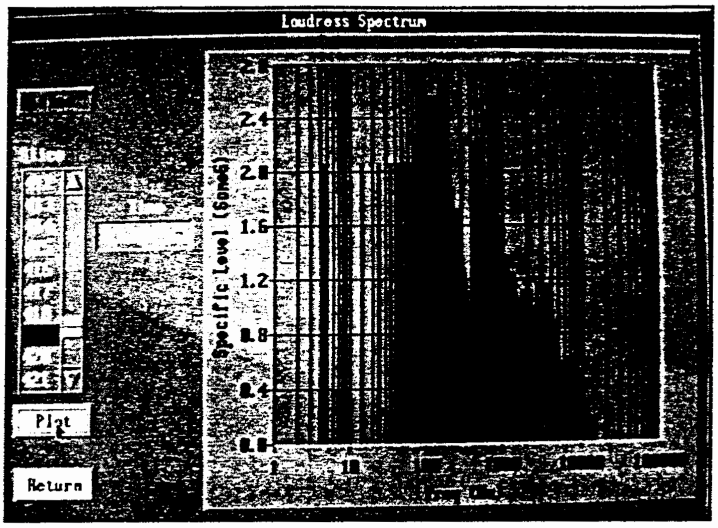
Roughness, like many metrics, can be considered independently of other metrics. Unlike fluctuation strength, roughness is perceived as a constant "rough" sound; while the listener can detect the quality of sound produced by variation, the modulation rate is too fast to actually hear the variation as alternative sound levels. Thus, roughness pertains to a sound perceived as rough but not varying. Roughness can occur in time variation or in frequency variation.

ROUGHNESS DEFINITION

Roughness, a modulation-based metric, begins at about 10 Hz modulation frequency and sees its maximum around 70 Hz modulation. In order to produce roughness, modulation must occur in the region between 15 and 300 Hz. The unit of roughness, the Asper, is referenced to a 1 kHz tone at a 100 percent modulation in amplitude and at a modulation frequency of 70 Hz. For amplitude modulation, the important parameters are the degree of modulation and the modulation frequency. For frequency modulation, the modulation frequency and the modulation index become the important issues. Typical narrative descriptors used with roughness include roughness and harshness.

FLUCTUATION STRENGTH

The sensation of modulation is heard when modulated sounds exhibit a modulation rate of up to 20 Hz. Above this modulation rate, the temporal perfor-



Loundness spectrum.

mance of the listener begins to saturate, and roughness sensation begins. This metric is able to be considered in the absence of other metrics. Fluctuation strength increases with increasing sound pressure level.

FLUCTUATION STRENGTH DEFINITION

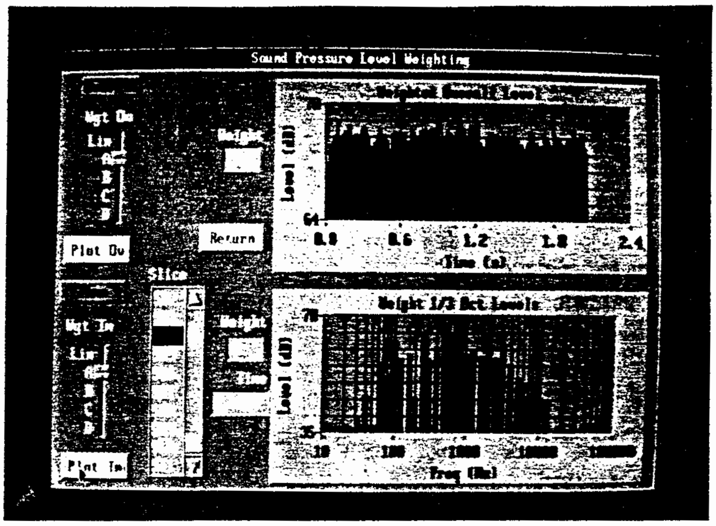
Fluctuation Strength is useful as an absolute and as a relative value, and fluctuation can occur in level or in frequency. It is affected by sound pressure level, by modulation depth, by modulation factor and by center frequency.

A fixed reference selected is a 60 dB 1 kHz tone 100 percent amplitude modulated at 4 Hz. This is referenced to 1 Vacil. For a given sound, the maximum fluctuation strength is found at a modulation depth of at least 30 dB. Above the maximum value, fluctuation strength remains constant.

A CALCULATIONAL EMBODIMENT

Having discussed the four principle metrics being used in sound quality calculation, there is clear interest on the part of many manufacturers in beginning to evaluate the benefit of these and other metrics to their sound quality and noise control programs. The only metric in common use at this time is loudness, as the use of the others has been substantially limited by the cost of software and testing systems needed for their use.

As part of the Organization of the Sound Quality Working Group, (See Sound & Communications, Sound Quality, Part 1, March 1992), Orfield Associates has just completed the development of a software package, Sound Quality Analysis I, which is the only inexpen-



Frequency weighting.

sive package available at this time for use in sound quality calculation.

This SQA I package, shown in examples within this article, will accept either direct input of microphone signals and will accept ASCII analyzer files of both single spectrum and multi-spectra data. In its first revision, it will read data directly from certain B&K analyzers.

Thus, measurements taken previously can be calculated in order to begin to develop correlations between different products or product prototypes, and current work. This software package is now in the hands of a number of acoustical departments of major manufacturers. Via the use of binaural recording and digital editing, jury presentations, and sound quality calculation, the product manufacturer can now begin to inexpensively consider the breadth of the sound quality field. In addition, by joining the Sound Quality User Group, he can be directly in touch with his professional peers in other firms.

SUMMARY

Via the consideration of sound quality calculation, the sound quality engineer or researcher can begin to develop a model for the analytical process of sound quality work. Via the use of listening juries and calculational metrics, the validity and predictive value of calculation can begin to become established, the specific metrics which are useful can be identified, and eventually, less listening jury work and more analytical work will begin to reduce the time burden and the expense of using listening juries as the only method of sound quality research.

This revolution in analysis will finally allow many manufacturers to begin the long process of converting from noise control engineering to sound quality engineering. In our next sound quality article, we will discuss alternatives for statistical consideration of SQ data along with alternative jury testing instrument methods available to this field. ■