SOUND (QUALITY) SENSE

Many of us would agree that there are a lot of things around us that are too noisy, such as the seats behind the engines on a long flight, an angry two-year-old and the volume of the music system in the car next to us at the stoplight, which shakes our seats. Sometimes, though, it's not the loudest sounds that are the hardest to live with. It's the intermittent hiss, or the whine perfectly pitched to set our teeth on edge.

When a product or an appliance is manufactured, a vehicle, a toy, a train carriage, anything that people are going to hear or be near, it should sound "right". Not just quiet enough to be unobjectionable **-** that's necessary but not sufficient. It's got to have that indefinable Sound of Quality, the sound you would wish your car door in closing would make.

A truism of sound quality work is that a good lawnmower doesn't sound like a good refrigerator. Companies who take this sort of thing seriously do jury testing. They may use listeners like their customers to get a better measure of public opinion, or listeners from among their engineers, to gain a better understanding of which elements of the product are making the most objectionable part of the noise. The jury ranks noises according to various rules. The 2 most successful forms of comparison are "Paired Comparison" and "Semantic Differential". The use of a Fixed Scale ("On a scale of 1-10, how annoying is this noise?") is generally not recommended.

Paired Comparison works by comparing 2 sounds, much like you do when getting an eye examination for glasses or contacts. Do you like sound A or B? If A is chosen, do you like A or C, etc. The better tests present the sounds in different orders to different juries. Semantic Differential compares the qualities of the sound. Does A sound more pleasant/annoying/sharp/etc. than B?

Since jury testing is expensive and time-consuming, it's important to get as much information from the testing as possible. If an analyzer offers metrics that correlate well with jury responses, it becomes easier to pinpoint the areas of the product that will benefit from improvement (at least until that part gets so good that something else starts to be heard as the problem instead). Note that no one knows in advance what metrics will correlate well with jury preferences for a given product so analyzer manufacturers offer an extensive array of tools to choose from. Which metric to use is something a company has to establish for itself, though some rules of thumb are going to become apparent as a company gains experience in sound quality testing. Jury testing will continue to be important and an analyzer that helps to smooth the process by for example, being able to play a series of sounds as recorded or as filtered in different ways, has an advantage

It is hard for measuring instruments that are designed to measure absolutes, to give a good indication of how people will feel about a sound and to provide a subjective measure. Still, with many juries listening to many sounds of all sorts, researchers have developed metrics that sometimes prove useful in predicting the response of human customers and their neighbors.

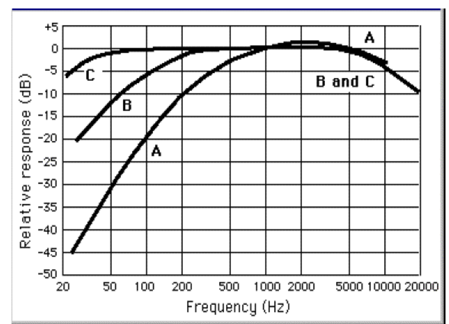
These metrics do not have any intrinsic correspondence to human response. Each of them is an approximation. Choosing among the available metrics requires some experience of the way that people respond to a particular product's sound. The ability to use sound quality software effectively in a specific product environment will evolve as the observer learns more about the appropriate measurements that help to form the best predictors of customer response, so it's important to choose an instrument that offers a well-filled toolbox**.**

A good Sound Quality instrumental system must not only provide the ability to make objective measurements of the highest accuracy, but also bring the same care to the more subjective measurements required for Sound Quality work. These must include all the usual objective measures such as Power Spectrum and nth Octave but also include a respectable set of subjective measurements. The subjective metrics that form the basis of a Sound Quality instrumental system include a broad assortment of tools for understanding the human appreciation of sound. With increasing experience in the use of sound quality measurements, an observer will be able to reduce reliance on jury testing with its attendant costs, delays, inconvenience and uncertainties. With reduced test time comes shorter time to market. Repeatable ***subjective*** measurements may offer the insights needed to tackle sound quality at the design stage instead of as a fix for a perceived problem in the consumer's hearing.

The specific objective measure used to determine sound intensity is called **Sound pressure level (SPL)** and is the straightforward measure of acoustic pressure, given in dB relative to 20μPa for measurements in air (measurements underwater traditionally are taken relative to 1×10^{-6} Pa). A, B, C or D weightings are applied to better represent the human hearing condition. **Octave** analysis and **Statistical measures** such as L_q & L_n. and other measures are also used.

One aspect that has to be taken into account is the way the ear responds to different sounds since the ear seems to respond more in proportional frequency terms than linear terms. Consequently octave analysis was developed, and it and its off springs have dominated acoustic analysis. However, octave analysis isn't completely satisfactory in representing the way ears work either. There are in effect, filters in the mechanism. For instance, when one sound is already present, sounds near to it in frequency will be less audible, a phenomenon called masking and the masking effect can be used to map out the frequency band which one might imagine represents the range or band over which the ear is non responsive. For instance, if you play a band of noise centered at 1kHz at 60dB SPL, you will find a tone at 1kHz needs to be 57dB before you can hear it, while a tone at 750 Hz or 2kHz could be heard at about 10dB. The area over which significant masking occurs outlines a critical band.

Standard filter contours A, B, C and D are used to make measurement instruments more nearly approximate the response of the normal human ear. The different contours are intended to match the ear's response at different sound intensities. When making practical assessments of the sound level of a concert for example or as a part of a general survey of ambient sound levels, the type of measurement that is usually made is that of the sound level in dBA. This measurement is made with a sound level meter or Sound & Vibration Analyser with an A contour filter which provides the best instrument match of the ear's equal loudness curves for soft sounds in the neighborhood of 40 dB. When this filter is used, the levels should be recorded as dBA rather than dB. Measurements made in dBA approximate the loudness level in phons. *(HyperPhysics Georgia State University.)*

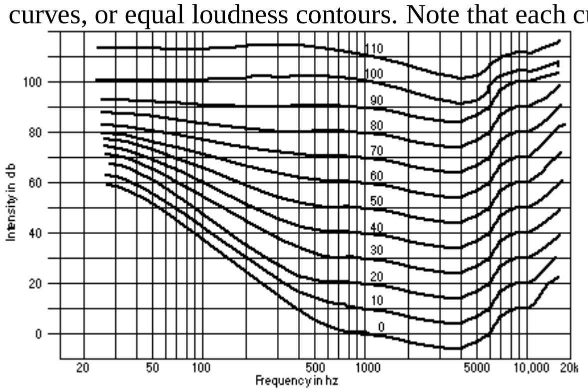


The A, B and C Sound Level Contours

The A-contour filters out significantly more bass than the others, and is designed to approximate the ear at around the 40 phon level. It is very useful for eliminating inaudible low frequencies. The intermediate B-contour approximates the ear for medium loud sounds and is rarely used. The C-contour does not filter out as much of the lows and highs as the other contours. It approximates the ear at very high sound levels and has been used for traffic noise surveys in noisy areas. The D contour is used to measure aircraft flyovers.

Loudness. Sound loudness is a subjective term describing the ear's perception of the strength of the sound. It is intimately related to sound intensity but can by no means be considered identical to intensity.

Often people will say one sound is louder than another even if both sounds have the same sound pressure level or measured intensity. Since the early part of the 20th century attempts have been made to 'objectively' rate subjective loudness. There have been many attempts to create a measurement that would quantify a person's perception of loudness. Most were designed for a particular kind of noise and correlate neither with each other nor with listeners for a broad range of sounds. Fletcher and Munson took a large group of people and had them listen to pure tones. They adjusted the volume until two tones at different frequencies sounded equally loud. The curves charting the connection between SPL, frequency, and loudness became known as the Fletcher-Munson curves, or equal loudness contours. Note that each curve is given a number equal to its sound pressure level at 1kHz



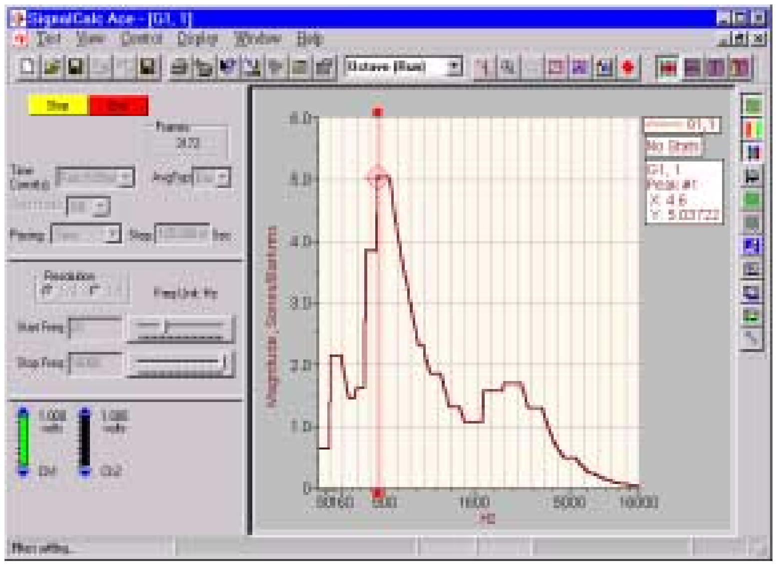
In most cases A-weighting is a satisfactory method for ranking noise in approximately the same way as it is subjectively heard. Work carried out by E Zwicker and Stevens with a view to an International Standard for loudness calculation resulted in ISO 532 (1975), albeit with some dissent. However the Zwicker method has come to be seen as the most useful method and is often used in sound quality instruments.

Fletcher-Munson curves

The Zwicker method is designed to be used with third octave measurements. Briefly the 'critical bands' or 'frequenzgruppen' are approximated by third octave bands, but for low frequencies by two or more third octave bands summed together. That is, total loudness is a summation of specific loudness across critical band rates with an additional factor for the effect of specific loudness on one band on adjacent bands. While in general critical bands are wide enough that masking is localized within them, if you picture each as a bandpass filter you can see that if the level in one critical band is much higher than its neighbors, even the skirts of the filter will have an effect and the next band will be masked.

The Zwicker method can be used for free field or diffuse field measurements giving results in units termed as sones GF or sones GD, and phons GF or phons GD. The Zwicker procedure for calculating loudness is quite complex as it involves plotting the band pressure levels on one of a series of ten charts and then measuring the area under the figure in order to derive the loudness. By using an instrumental measurement system based on a computer, the software greatly facilitates the calculation procedure and provides graphical output. Acquisition and measurements of the metrics of sound is automated, graphed and recorded. The Data Physics SignalCalc range of Sound Quality and Dynamic Signal Analysers offers such facilities.

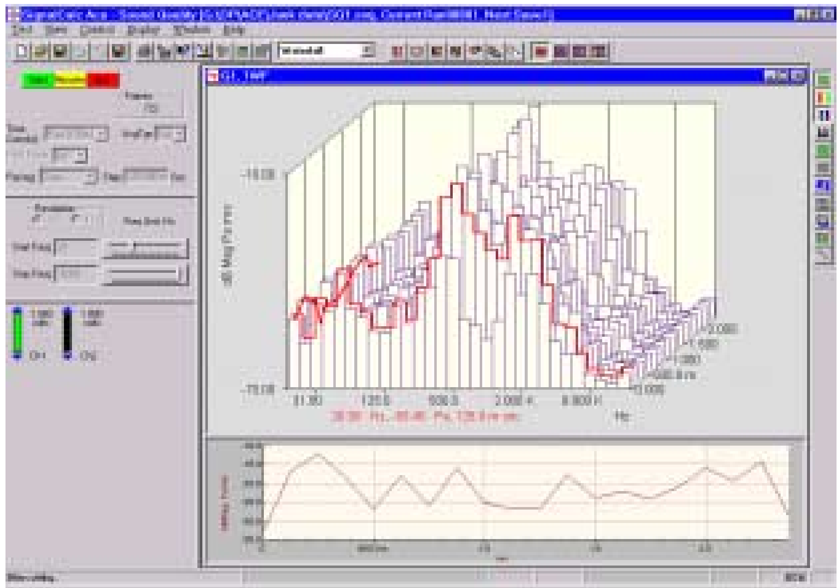
Loudness shown in Sones / Bark rms.



Measurement of loudness in phon has two main drawbacks. One is that a person tends to report that a 10 phon increase in level sounds twice as loud, so the increases are not intuitive. The other is that the correlation of phon with perceived loudness is not good for anything but sine waves. To correct the first problem, the sone was developed as a yardstick for which twice as many sones indicated a person heard twice as loud a sound. One sone is the level produced by a 1kHz tone of 40dB SPL. Initially, the conversion between phon and sones was simply a matter of a dB-to-linear calculation, with some scaling to end up with nice numbers, but again that only took care of the sine waves. Later, however, researchers spent considerable time and effort on developing rules for measuring loudness that would match up with average human perceptions for a wide range of sounds, and a variety of ways to measure loudness in sones appeared, some of which correlate with each other scarcely at all.

Time varying loudness gives loudness as a function of time and as filtered in a way that approximates the ear's response to sound. This can be total loudness or specific loudness as a function of critical band rate, in which case it is shown as a 3-D display. A waterfall displays time-varying specific loudness in sones/Bark, which you may slice to show specific loudness as a function of frequency at a selected time (record plot), or specific loudness as a function of time for a selected frequency band (slice plot). You can also select Waterfall functions to show total loudness or sharpness as a function of time in the 2-D display. Other sound quality metrics are shown as annotations. The **statistical distribution** of loudness is reported as Nn, where N is the loudness (in sones relative to a 1kHz pure tone at 40dB SPL) that is exceeded n % of the time. N_5 is often used as an indicator for the perceived loudness of a strongly varying or intermittent source. Software that offers waterfall displays of time-varying loudness in sones and which can be sliced in x and y provide an infinite number of analysis views.

Sharpness is another useful metric, measured in **acum**, which reflects the high-frequency content of a sound relative to a narrow-band noise. The reference value of 1 acum is defined as a narrow band noise, one critical band wide at 1kHz at a level of 60dB. Sharpness is calculated by multiplying the specific loudness function by a weighting function that increases drastically above 16 Bark then by summing and normalizing by total (unweighted) loudness. Time varying sharpness is shown as a 2-D plot or waterfall format, while statistical distribution of sharpness S_n is reported alongside N_n as an annotation.



A Waterfall plot shows magnitude in Pascals rms and one slice plot.

Time variations of low frequency modulation on a shorter scale, (up to about 20Hz) lead to sensations of **Fluctuation Strength**, which is measured in **vacil**. The ear seems particularly sensitive to 4Hz fluctuations, no doubt because that is approximately the frequency of syllables in speech. Unfortunately, if there is no useful information in those variations, they're perceived as annoying. One **vacil** is defined as the fluctuation strength produced by a 60dB, 1kHz tone 100% amplitude-modulated at 4Hz.

More rapid modulation (15-300Hz) leads to a measure of **Roughness**, measured in **asper**. One **asper** is produced by that 60dB 1kHz tone if it is 100% amplitude modulated at 70Hz. The sensation of roughness peaks at around 70Hz. The region from 15 – 20Hz is something of a transition band and has some effect on both fluctuation strength and roughness. There is also debate over the proper upper limit for measuring roughness. Everything from 100 Hz to 300 Hz can be found in the literature, mainly because the ear becomes much less sensitive to roughness as the frequency increases and begins to perceive a pitch in the variations instead. It is more usual to expect a roll off as the frequency gets above 100Hz rather than to see a sharp cutoff.

Fluctuation strength and roughness take into account summations over frequency and time and are therefore not reported as a function of time but as a number associated with a measurement or suite of measurements.

Unbiased annoyance incorporates loudness, sharpness, fluctuation strength and roughness all into one metric, which enables the description of the annoyance of a sound with some expectation that a listener will agree with you. This metric is known as **Psycho-acoustic Annoyance** and attempts to measure the annoyance of a sound in a way that correlates with the experience of a listener. Zwicker's equation for Psycho-acoustic Annoyance, also called unbiased annoyance by some, is the one most often used to determine Psycho-acoustic Annoyance.

Computer based instrumental systems are the most suitable products for these measurements as they can give rise to many graphs and diagrams to enable easy visualisation of the results of the measurements. Graphs which can show loudness and sharpness viewed in spectral density terms, overall, or over time with statistical distribution Nn, Real Time Spectrum, 1/3 Octave or Waterfall graphic display and showing parametric annotation, measurement readout, peak cursors and plotted harmonic values.

There are many other metrics which are used in specific cases such as sound power, sound intensity, reverberation time (T60), prominence ratio, tone-to-noise ratio and RASTI but these measures are outside the scope of this paper.