Proceedings of the Institute of Acoustics & Belgium Acoustical Society
Noise in the Built Environment, Ghent, 29-30 April 2010

RECENT ADVANCES IN SOUND QUALITY

P Van de Ponseele LMS International NV, Leuven, Belgium

1 INTRODUCTION

Sound Quality is not really about music or harmony. Most think of Product Sound Quality when discussing Sound Quality. The sound of the product should match the expectations of the user of that product, in a particular context. The automotive industry is well aware of this and invested heavily in this topic during the last twenty years. With the venue of electrical engines, a new set of challenges needs to be addressed. The silent electrical motors do not bring sound quality without effort – there is more to Sound Quality than just lower levels.

By definition, Sound Quality is subjective as perception differs for each individual. General trends on the perception of Product Sound Quality can be observed and agreement often reached, especially on extreme cases. Jury testing techniques including basic and advanced statistical processing are used to turn the subjective opinions into more objective ones. Jury tests are time-consuming and therefore costly. Solutions to problems cannot wait; on the other hand, relying on the opinion of experts working on the product proved to result in over-engineering and additional costs. Faster methods are needed to estimate objectively the sound quality performance of the alternative solutions. For this purpose, sound quality metrics, and even tools to develop these, are available and commonly used.

The A-weighted sound pressure level is the reference sound metric used for regulation and by international standards. Only for wider product ranges will this parameter correlate well with perceived product sound quality. For competitive analysis, Sound quality metrics are available with either a focus on level, annoyance, tonality, speech disturbance, and modulation. For specific products or operational conditions, combinations of these metrics are defined.

The tool suppliers for product design and development invested in extending their level prediction tools with time domain methods in support of listening, jury testing, and metric calculations. Although the quality of the sound prediction is limited, its benefit to just the levels is regularly huge.

In this presentation, reflections on Product Sound Quality will be presented based on experience in automotive and white goods industry including the process of making perception more objective and tuned to a target market.

2 PRODUCT SOUND QUALITY, A SCIENCE OR …

Psychoacoustics is the science of how we hear. Our hearing is determined by geometrical parameters of our head and outer- and inner ear. These dimensions are different for all individuals. The variance resulting from these differences is not negligible but can be analyzed and largely be understood by physicians.

Our perception is influenced by how these acoustic stimuli are processed in the brains; this influence depends strongly on the individual. The processing of the acoustical stimuli happens in an environment where multiple stimuli of different kinds (visual, motion, sensory, etc) need processing in parallel, also taking into account that the individual can be preconditioned emotionally and physically prior to all of this. Angry people will react more vividly to sounds when these are directly related to what or who is making them angry. Tired people might also be more sensitive to the noise of playing children, while those taking a ride on a roller coaster hardly experience the sound as a relevant parameter of the ride.

The analysis of hearing perception is very hard; each individual seems to be a time-variant system of a population with a huge variance to start with.

Working in an environment of engineers in applied sciences, learned that none have issues with level or frequency based acoustics even if the variance on the average is 'too big to be good'. However, once hearing perception is discussed in terms of adjectives or emotions, many feel uncomfortable, some even state that this is no longer science ... just 'marketing'. Questions like "How can a sound be expensive or comfortable?" are commonly heard.

Still, all of us have been impacted emotionally by music at some time. All of us perceive an increase in security or comfort from the acoustical feedback of devices that we use daily. We all know the sequence of sounds of our coffee machine. The sounds of an elevator confirming our call and coming our way, the sound of the car of a beloved one arriving on the driveway just as we are expecting him or her, the sound of the engine of your older car when starting on a very cold day, all these sounds give us a good feeling.

The above examples of every day sounds indicate that the sound should match the expectation of the product or event. The sound for the coffee machine should match our expectation for that sound; we are waiting for the sounds, any delay makes us worry. The same applies for the car arriving on the driveway. The sound of the starting engine will make us feel relieved but we might not like that loud sound at all, a few minutes later when driving on the highway.

A scientific approach to this type of problems requires a proper gathering of perception feedback as input for statistical processing. The complex perceptions need to be converted to numbers that allow design choices to products to be made objectively. Jury testing is a suite of techniques appropriate for Product Sound Quality design. The food industry pioneered this field.

3 JURY TESTING, A SCIENCE ON ITS OWN

Jury tests can be performed using the real test object in its true environment. There are hardly issues with the recreation of the total perception environment. However, the visual and tactile feedback of the different test objects creates an enormous brand influence bias. Testing vacuum cleaners or cars are typical examples. Driving the car of your personal preference affects your scores consciously and subconsciously. Acoustical jury tests of this type are impractical most of the time. The sounds are limited to what physically exists and the cost is high.

Most of the time, tests are executed in a controlled environment where brand bias can be controlled and manipulated sounds can be added to the set of test stimuli. The test sounds are called the stimuli and the people performing the tests the subjects. The so-called naïve subjects are people that are just users, or potential users, of the product and not technical experts in the domain of the test object producing the sounds.

These listening tests suffer from an over-attention to the sound stimuli as other potentially perception interfering stimuli (visual, sensory) are minimized, or kept constant.

Many aspects to jury testing require special attention and mistakes are made easily. Some mistakes are obvious others painfully gained by experience. The incomplete list below gives an overview of most of them.

- Phrasing of the question
- Selection of the subjects (people involved in jury test)
- Selection of the stimuli (sounds used in the jury test)
- Room listening or headphone listening
- Context available or not
- Selection of jury test method
- Selection of relevant subject properties

- Preliminary hearing test or not
- Size of test
- Multicultural or local

These aspects will be discussed one by one in the sections below. It is clear that all aspects need to be taken care of for a good jury test.

3.1 Phrasing of the question

The questions in the jury tests are easily suggestive or insufficiently exact, especially when using adjectives. One can ask subjects to rate the variability of the perceived level, or ask for the irregularity of the sound. Irregularity has a negative connotation and is not exactly the same as a varying level; varying frequency content can be perceived as irregular too.

In the process of defining a technical metric for irregularity of diesel engine sounds at idling, the naïve subjects were asked to rate the sounds for irregularity. A couple of sounds were twice in the test just with a difference in level. Our subjects rated the louder sounds as more irregular while there was no reason to do so. It is clear that explaining what the question is about, in this case irregularity, without conditioning the group of subjects in a particular direction, is crucial.

3.2 Selection of the subjects

The subjects selected for the test should be relevant users of the product and be 'naïve'. Although this is absolutely clear to all, this is very often not the case. With the excuse to do a quick jury test, just to confirm the most recent great findings, very often a number of people of the work floor are called in, and added to the team working on the topic. The results of these jury tests are highly predictive and have proven to lead to over-engineering of the product. The experts focus on technical/acoustical challenges and have lost a balanced perception for this product. The use of students is the other typical example – the excuse to use students is cost and availability. When using students to evaluate the sounds of luxury cars, it might well be that the more sportive sounds get overrated.

Four people in the irregularity jury test did not make the mistake of taking loudness for irregularity; they were the four people that prepared and organized the test. They could not be referred to as naïve subjects. People involved in the project being tested, tend to answer differently than the rest.

The most common error is made with respect to the selection of the number of subjects. Major decisions have been made and 'scientific conclusions' drawn from jury tests with just a handful of subjects. The minimal number of subjects depends on the several parameters but 25 is a minimum.

3.3 Selection of the stimuli

Binaural recordings should be used for jury tests. This is not a hard requirement when performing jury tests in a listening room, but when using headphones this is a must. Monaural recordings result in a head centered perception, which is not natural.

In some cases, the level of the stimuli is part of the evaluation by the subject. In most real life cases, small level variations are not relevant to the overall perception of the product sound. The perception of subjects in a listening test however, is very sensitive to the level as the sounds are presented shortly after each other. For this reason, loudness corrections to all sounds are often applied.

In a listening jury test on sportive sounds, a small Italian car came unexpectedly out as one of the best. Needless to mention the embarrassment for those involved in setting up the experiment. The reasons for this became clear after some analysis. The sound used for that small car was taken from a recording in second gear in order to keep the sound stimulus short enough. Through this operation, it became the shortest sound of them all. To minimize level bias, all sounds had been rescaled to the same average loudness. The rapid engine speed variation and the clearly audible engine orders made it sound more sportive and powerful than most other sounds in the test. Playing the sounds with the correct level and all sounds selected for the same speed range gave the expected results.

3.4 Room listening or headphone listening

The advantage of listening in a room is that our head and torso are also 'listening'. It is known that the vibrations generated by the impacting noise influence our listening. It is not easy and quite costly to create a sufficiently large sweet spot for accurate listening with multiple subjects.

Movements of the head are perceived as natural; this supports a more realistic sound perception.

Listening with headphones can be organized with closed headphones or with open headphones, the latter with or without support of a sub-woofer. Testing with open headphones requires a listening room with a low noise floor. In all cases, equalization- of the stimuli and headphone corrections should be applied. The fact that subjects are wearing a headphone during the tests is an influencing parameter with a hard to predict impact.

In the case of room listening, some spectral shaping is needed to get a 'flat' room response. Listening rooms have quite some absorption; some are almost semi-anechoic chambers.

While gaining experience on jury testing and writing software for this, open headphones were used in an office environment and the subjects were seated in front of workstations. It was noticed that the scores obtained for a particular workstation were different from the other ones. One of the subjects mentioned in his comments that there was a steady, low-level, annoying tone coming from the left. The root cause was one of the disk drives of that workstation combined with the subject's location seated parallel to a window. Strangely enough, this tone was not audible when the headphone was taken off.

3.5 Context available or not

Listening jury tests require a minimum context to get all subjects focused on the same topic. Giving context for each stimulus creates bias most of the time. When the stimuli are no longer called 'A' and 'B', but are named after the brand of the car, results will differ. Typically, a fixed image or projection, and a spoken or written introduction are used to describe the context.

Jury tests for automotive applications sometimes prefer to do the test in a car simulator. Several subjects lost their focus on the test and 'scanned' the features of the test vehicle/setup. This setup provided too much context for them.

Several years ago, a set of eight binaural recordings of very different cars circulated in the company. This set had two versions: the original version with the correct names of the cars and a copy where the cars sounds were numbered from one to eight. When I was assigned to the Sound Quality software development team, I decided to do a simple ranking test using the second set with the numbers; at that time, I was not aware of the existence of the original set. When a colleague provided the original set, I was surprised about one particular rating of a very fancy car that I had put on 5^{th} place. I listened again, and yes, I had come to reason and this fabulous vehicle was now second. Only days later I realized how I had cheated myself.

3.6 Selection of jury test method

The choice of the jury test method is determined by what type of result is aimed at and with what level of quality. Below the commonly used methods are described and explained.

3.6.1 Ranking test

A relative small number of stimuli are presented to the subject and this subject has to rank them from highest score to lowest score. During the test, the stimuli are switched to new positions and quickly the subject looses the reference to the original order. On request, the stimuli of choice are replayed and the subject takes the time needed to rank them.

This test has a simple setup and provides most of the time a clear winner. Relative information is only indirectly and not accurately available at the end. Naïve subjects regularly do not use the same criterion for ranking the stimuli, especially if the criterion is an adjective.

Ranking tests are suited for a quick validation by a group of experts.

3.6.2 Rating test

The subject has to give a rating for each stimulus. This rating can be a number or a position on a slider bar. This method is suited for larger sets of stimuli. For proper quality of the results, a training session prior to the actual test is needed to avoid subjects to rate within a narrow span while waiting for the perfect or worst possible sound. In order to monitor drift in the ratings and check the subject's consistency, the stimuli need to be repeated at least once.

This relative fast test will benefit from statistical processing by experts. Corrections for individual rating spans, crosschecking the ranking, and removing subjects from the analysis based on consistency doubts, are actions that might be required.

Rating tests provide quickly a lot of information but regularly end with discussions on the results.

3.6.3 Semantic difference test

This type of test is similar to the rating test except that a bipolar pair of adjectives is presented. Typically a number of bipolar pairs are presented per stimulus and the subject has to rate all of them prior to going to the next stimulus. This technique is commonly used to identify relations between adjective pairs and the positioning of products with respect to dominant axes determined by factor analysis.

Soft	Hard
Powerful	Unsatisfactory
Quiet	Clamorous
Clear	Thick
Strong	Weak
Expensive	Cheap
Smooth	Rough
Thick	Thin
Ringing	Booming
Relaxed	Tense
Wide	Tight
Crisp	Not crisp
calm	Shrill

Fig. 1: Example of pair of attributes by Murata et al[1]

In the publication of Murata et al[1], three dominant factors were determined: comfortable, powerful, and booming. The first factor comfortable correlated well with soft, quiet, expensive, smooth, relaxed, wide and calm. The second factor powerful, correlated well with powerful, strong, and thick. The individual stimuli (read test objects) could then be visualized in that three-dimensional space.

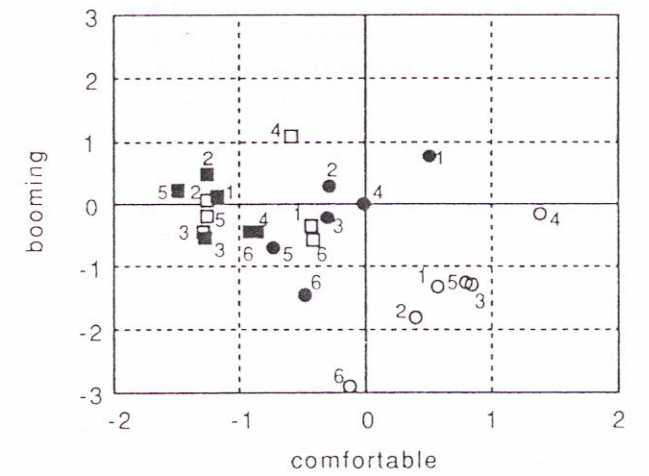

Figure 2: Evaluation space of comfortable and booming factors[1]

3.6.4 Paired comparison test

The stimuli are presented as pairs to the subject who has to select one of them. In theory all possible pairs have to be presented in both orders. Dedicated algorithms are used to space optimize the spacing in time of the stimuli[2,3]. Ten (10) stimuli will result in ninety pairs (90=10*(10-1)) to evaluate. When the stimuli are perceived identical in perception, the hard choice to make generates quite some stress. To avoid building too much stress, sometimes a gradient in preference is used.

Example:

- 'A' is preferred
- 'A' is just preferred
- No preference
- 'B' is just preferred
- 'B' is preferred

The A/B paired comparison technique, using the exclusive choice 'A' or 'B', has proven to be the most reliable technique from a statistical point of view. It is a time consuming technique, especially with respect to the number of stimuli in the test. Typical paired comparison tests have only six, seven, or eight sound samples. The figure below gives the overview score table for a test with 25 subjects. The test included the replication meaning that all pairs were presented in A-B and B-A mode.

Scores	A	B	C	D	E	E	G
A	-	2	4	46	48	10	42
B	48	-	44	48	50	50	48
C	46	6	-	46	46	18	48
D	4	2	4	-	34	6	44
E	2	0	4	16	-	2	46
F	40	0	32	44	48	-	50
G	8	2	2	6	4	0	-
Totals	148	12	90	206	230	86	278

Figure 3: Example of overall score table for paired comparison test

Relative ratings can be derived from the scores using the variations of the preferences of the individuals.

3.6.5 Difference test

This test is an alternative to the paired comparison test except that the question to answer is on how different the stimuli are perceived for a given parameter. The possible answers are in an enumerated form: from 'Equal' to 'Very Different'. Often the difference question is not unique and several aspects of the perception of the product are included (pleasant, powerful, metallic, etc). These tests are used when the goal is to find a niche for its product sound. Dedicated statistical processing is needed to process the results to positions in a multi-dimensional space. Another application is to determine how relevant proposed changes are to subjects. Other jury tests will indicate that some sounds are better than the current product sound, but are these improvements relevant.

3.6.6 Other jury test methods

Combinations of the above tests exist as well in the execution of the test or the statistical methods used with them. A rating test can be converted to a paired comparison test when taking a focus on the differences of consecutive ratings. This method is also called the Latin square.

3.7 Selection of relevant subject properties

There is the protection of privacy and not all relevant questions might be allowed. In general, it is good to ask a few questions to the subject on his experience with the test object, these questions should have simple answers to keep the stress level low prior to the actual test.

One should not forget to ask feedback at the end of the jury test session. Very often, the subjects feel the need to express something and should be able to give this feedback freely at the end. This feedback is useful to improve jury test setups. It is an indication of respect of the organizer towards the subject … the organizer might need the subject later again.

3.8 Preliminary hearing test or not

A minimal hearing test is very worthwhile but hardly ever done. The expected gain for the organizer seems low compared to spending the time for the test. How the organizer should deal with imperfect hearing is also not straightforward. Larger companies with a database of subjects make the effort.

A classical excuse is that the persons with hearing issues will be removed from the statistics at the end as they will have exceptional scores – this might be the case sometimes but not al the times. Removing subjects from a jury test data should not be done lightly.

3.9 Size of test

When starting with jury tests, organizers typically exaggerate with the number of stimuli or the overall duration of the test. After about 45 minutes of making tough decisions, most subjects are tired. Giving subjects control over the pace of the test, which is not possible for group testing, will allow subjects to continue longer, never to exceed an hour and a half.

The number of subjects should equal and preferably exceed 25 people as mentioned above.

3.10 Multicultural or local

It is clear that sound perception and perception in general depends on our culture. Some regions appreciate powerful sounds from the engines of their cars; others search for a sportive sound and what this means depends on the local taste. Jury tests should be executed with an audience of the local market where the product is aimed for. This is common procedure in the automotive world.

In the context of the European' research project OBELICS, an attempt was made to provide the bipolar pairs of adjectives in English, German, French and Italian. Most translations were straightforward but even after several iterations some adjectives did not have a meaningful equivalent in one of the languages.

In automotive, it is commonly known that a sportive sound for a Japanese public should contain higher pitch tones than for the European market.

3.11 Summary on jury testing methods

Jury testing enables ranking, rating, and positioning product sounds. It requires experienced persons to organize the tests and interpret the results. It turns a lot of subjective information into an objective set of results, being aware of the large spread on the results and the limitations of the test environment.

The results can either be used for trouble shooting, component selection, product marketing, and for future product design.

In general, the results of jury tests are correlated with sound quality metrics calculated on the stimuli. The metrics that correlate most with the subjective scores are then used as quick evaluation tools to monitor the progress made on the evolution of the product sound. A final jury test to confirm the findings is recommended.

The above-mentioned jury testing methods are passive; the subject listens and logs the score. Involving the subject in simple actions such as driving on a straight road or asking to perform a particular maneuver in a simulator, reduces the focus on the sound[7]. The simulator environment cannot guarantee correct focus and the subjects might all have performed the maneuver in a different way resulting in additional variance on the scores.

4 PRODUCT POSITIONING

Sometimes it is not enough to have a good product from a sound quality point of view; it should also be recognizable, specific to the brand and valued by the potential buyers. Most high-end automotive companies have selected their spot in the multi-dimensional environment of perception. For high-end cars, having a brand sound is almost a requirement. For these products, more technical and financial means for developing a brand sound are available. The brand sound is only one of the specific brand properties, but an important one as it is able to convey messages of many kinds.

For the family-car market segment, the challenge for the car manufacturer is to select a property as a brand item. This property is not commonly sound, this property can be safety/security, handling/fun, robustness/good-investment, environmental/Kyoto, or other ones. It is for them important to be recognized for their selected brand item and not to fail on the other properties.

To position a sound in the multi-dimensional space of perception, for a car in a specific mode of operation, a set of properties is chosen. The most common ones are related to the perceptions of sportiveness and comfort. Other commonly used ones are related to lively, powerful, expensive, and pleasant. Bipolar pairs are commonly correlated with a number of more basic pairs, it is as such useful to know to what extent and to which 'basic pairs' this perception is correlated, but once this is understood, the advantage of the direct questions is the ease of interpretation of the results.

Multi-Dimensional Scaling is a technique used to identify the location of a sound in a multidimensional space. The technique can be easily explained as if one would be asked to generate a map with the location of a number of cities just knowing the relative distances between all of them.

The MDS experiment uses as input the results of difference tests or the input of rating or bipolar pair tests where the 'absolute' ratings are converted to relative ratings. The set of stimuli should include the competitive products and might include some of the options provided by the engineers and retained by product management.

Below you find a typical, simplified example.

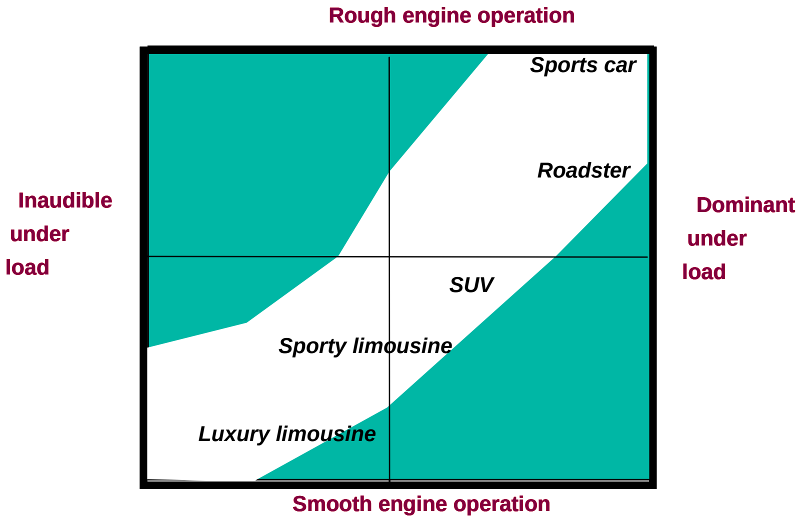

Figure 4: Example of product positioning of cars

5 SOUND QUALITY METRICS

The sound quality metrics can be divided in several groups of metrics. A typical system to split or group them uses six groups. These groups are:

- Level metrics
- Modulation metrics
- Tonal metrics
- Annoyance metrics
- Speech metrics
- Application specific or technical metrics

5.1 Level metrics

The level metrics are the A- and B-weighted Sound Pressure Level, the Loudness according ISO532A, ISO532B, Stevens VI, or Stevens VII, and Time Varying Loudness according draft DIN 45631/A1. These metrics are calculated from a microphone recording. The goal is to calculate an objective level for a given sound recording. In principle, there is no appreciation of the sound included only the physical elements of perception.

These metrics are commonly calculated in the time domain but for pseudo-stationary sounds, frequency domain methods are in use too. Time varying loudness is typically used for short transient sounds only.

5.2 Modulation metrics

The variation in amplitude of the envelope of a sound is perceived by humans in a specific way. Low frequency variations are perceived as oscillations in level. Mid-frequency variations are perceived as harshness, while high frequency variations are not perceived in a particular way; these are perceived as two separate tones. Fluctuation Strength is a metric calculating the degree of low-frequency amplitude and Roughness is a metric calculating the degree of 'harshness' of the signal.

Fluctuation strength is in exceptional cases perceived as a form of power but in most cases as lack of quality. Fluctuation strength is commonly used as a metric in the domain of speech.

Roughness remains a hard-to-interpret metric. There are cases where roughness related well with a sportive sensation but most of the time it is a sign of problems. Signals with the same roughness value can sound very differently and be rated very differently too. It is a common procedure today to check for modulations in sound signals as a preparation step to find potential acoustical issues.

5.3 Tonal metrics

Pitch, Tonality, Tone-to-Noise Ratio and Prominence Ratio are metrics for quantifying the perception of tonal components in complex sounds. The white goods industry is driving the developments in this area. The high pitch tones of some electric and electronic devices are annoying and fatiguing the user.

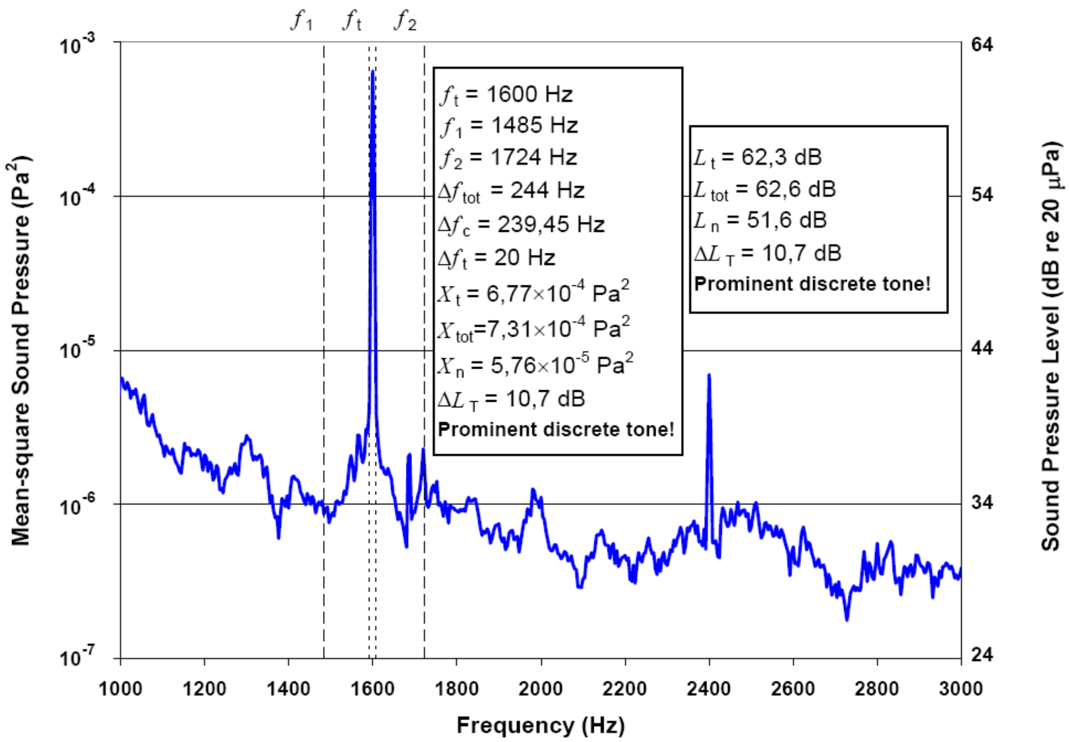

Figure 5: Example of tone-to-noise ratio method[4]

The automotive industry is getting increasingly interested in these metrics with the venue of the many small electrical engines in cars and electrical vehicles in general. In order for a product to be acceptable, very often the design threshold of a product for tonal metrics is very low. Perceivable changes of pitch are also not appreciated.

5.4 Annoyance metrics

These metrics aim at quantifying the degree of disturbance generated by environmental noise in working environments or buildings. Noise Rating, Noise Criterion, Balanced Noise Criterion, and Sharpness are commonly used. Sharpness is loudness based the others sound pressure level based. The perception of annoyance depends heavily on the expectations of the individual in the sound environment. The sounds of an elevator can be perceived quite differently by the person waiting for it to arrive, or the person living in the apartment next to the elevator. The same applies for the person accelerating with a sports bike through a narrow street and all standing by.

5.5 Speech metrics

Two types of speech metrics are commonly used that do not include speech itself. A first family of metrics takes a focus on speech privacy; a second family takes a focus on speech transfer. The speech intelligibility is quantified by evaluating the sound measured in its environment. Speech Interference Level (SIL), Preferred Speech Interference Level (PSIL), and SIL3 are metrics of the first family using the levels of three or four octave bands. From these levels, the quality and limits of speech are derived as a function of the distance between the talkers. Articulation Index is a commonly used metric of the second family. This metric uses the third octave bands from 200Hz to 4 kHz and allows estimating the percent of correctly understood syllables as a function of the articulation index and the type of speech.

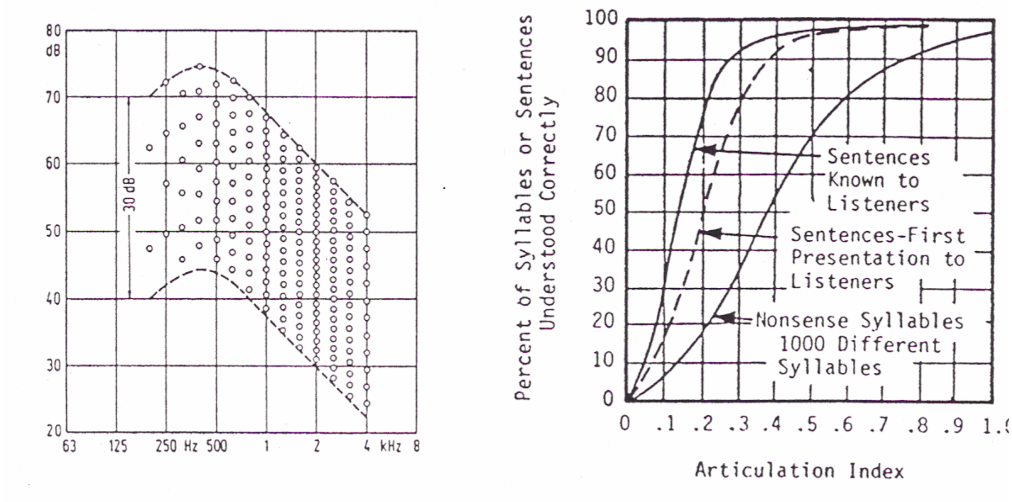

Figure 6: Articulation Index and the relation to speech understanding

The research in the domain of hearing aids pushes the developments for speech metrics. Speech Transmission Index (STI) and Rapid Speech Transmission Index (RASTI) are standardized for predicting intelligibility but improved metrics have been presented recently[5]. STI uses a signal that contains amplitude modulation characteristics similar to natural speech. The intelligibility is reduced by echoes or noise 'drowning' the speech itself. The result is a rating between zero and one. From 0.6, the intelligibility is good to excellent.

5.6 Product specific sound quality metrics

The correlation of jury test results with loudness, the A-weighted sound pressure level, and the articulation index is in many cases very good but not suited to distinguish better product sounds from the other ones.

For this purpose, engineers develop dedicated metrics that are a combination of regular metrics. The equation itself either is the results of solving an over-determined set of equations or based on engineering based trial and error, or most of the time, a good mix of both.

These technical metrics are valid for a specific group of products and a given condition. Idling noise of four stroke four cylinder engines could be the subject of such a metric. These metrics cannot be used outside their scope. Quite often, these metrics get descriptive names like Rumble, Muddiness, Harmonic Gradient, Irregularity, and Annoyance Index[6].

6 CONCLUSIONS

Good Product Sound Quality is about meeting expectations of customers. This paper describes the tools and methods that are available in support of turning the subjective opinions of many into objective ratings and conclusions.

Customers expect quality and relevance of the messages generated by the products in our environment. Personalized ring tones of mobile phones giving information on who is calling as a simple example. Other products should just not be audible; there is little feedback in tonal components of disk drives or ventilation fan noise. On the other hand, ventilation noise free of tonal components is desired in many workspaces in support of speech privacy and concentration of individuals.

Companies should also be aware of Product Positioning and the possibilities it provides. An easily recognized Product Sound can be a powerful sales asset and be in support of the marketing messages for the product.

Product Sound Quality is a complex matter and needs jury-testing tools to make better products. PSQ started in the automotive industry; but is used today for many other product fields. Research on improving hearing aids is a fine example.

Work is ongoing to turn listening jury tests into more interactive tests, with the goal to increase the relevance of the test environment and to reduce the primary focus on sound during the test.

7 REFERENCES

1. H. Murata, H. Tanaka, H. Takada, Y. Ahsasa, Sound Quality Evaluation of Passenger Vehicle Interior Noise, Proc. 11th ICA, Vol. 7, 85-88. Paris (1983)
2. H. A. David, The Method of Paired Comparison, Griffin's Statistical Monographs and Courses, Charles Griffin and Company, Ltd., (London 1969)
3. R. T. Ross, Optimum Orders for the Presentation of Pairs in the Method of Paired Comparisons, Journal of Educational Psychology, 25, (1934)
4. ECMA-74, Measurement of Airborne Noise Emitted by Information Technology and Telecommunications Equipment, 10th Edition (December 2008)
5. NAG/DAGA 2009, International Conference on Acoustics, Rotterdam (March 2009)
6. M. Hussain, J. Gölles, A. Ronacher, H. Schiffbänker, Statistical Evaluation of Annoyance Index for Engine Noise Recordings, Proceedings of the 1991 Noise & Vibration Conference, SAE International P-224 911080, (1991)
7. P. Van de Ponseele, M. Adams, K. Janssens, L. Vallejos, Virtual Car Sound Environment for Interactive, Real-time Sound Quality Evaluation, Proceedings of Inter-noise 2002, Institute of Noise Control Engineering, (2002)