Using the Metric Project in ArtemiS SUITE

In most cases, the quality of a sound depends on more than one parameter (such as the sound pressure level). Calculation rules that combine a number of different partial values to derive a characteristic single value result allow you to create a meaningful quality index for your sounds. Such calculation rules, also referred to as metrics, help you to assess the acoustic quality of your products in a reliable and time-saving manner, to determine strengths and weaknesses of your products, and to specify target sounds reliably.

Metrics can be developed, for example, based on the results of jury tests by mapping the judgments of subjects to measurement-based analysis results. With a metric developed that way, you can then reliably assess the perceived sound quality of your products in a time-saving manner that does not require additional jury tests.

The process of developing a sound quality metric involves several steps, for which HEAD acoustics provides you a range of tools:

- **Aurally accurate recording**: You can use, for example, the binaural artificial head measurement system HMS to create aurally accurate recordings of your sounds. The artificial head measurement system correctly models all acoustically relevant parts of the human outer ear, thus achieving an aurally accurate documentation of sound situations.
- **Aurally accurate playback**: A *lab*P2 unit allows you to play back a binaural recording with the correct signal level and equalization, thus ensuring a valid auditory assessment of your product sounds.
- **Perceptive assessment**: The SQuare software allows you to conduct systematic jury tests, and provides you a straightforward overview of the results as well as a statistical analysis of the sound judgements.
- **Measurement-based analysis**: ArtemiS SUITE analysis software offers you a wide range of analysis functions. Besides common standardized methods, such as level calculation, octave analysis, and the calculation of psychoacoustic parameters, it also provides unique analysis methods, such as the Relative Approach.
- **Metrics definition**: The Metric Project of ArtemiS SUITE determines the correlation between the perceptive judgements from a jury test and the single-value results of measurement-based analyses. The combination of multiple partial values with different weightings (single-value results of different measurement-based analyses) results in a robust calculation rule you can use for automatic assessment of your sounds.

By means of an example, this Application Note[1] explains how to use the Metric Project to specify a metric and how to apply the created metric. Additional Application Notes, e.g., about binaural recording and playback or psychoacoustic analysis, can be found on our website:
https://www.head-acoustics.de/eng/nvh_application_notes.htm.

---

[1] The descriptions in this Application Note refer to version 9.2 of ArtemiS SUITE. The general proceeding also applies to other versions. However, the scope of functions and the user interface may differ.

Description of the application example

In the following section, the various configuration options of the Metric Project are explained by means of an application example. For a better overview, this example has been reduced to a small set of only six sounds. In reality, you would need a considerably larger data set in order to create a reliable and meaningful metric. Important notes on how to achieve robust metrics can be found later in this Application Note.

For the application example, a jury test was conducted using the SQuare evaluation software. The six noise samples were judged by the subjects regarding their degree of annoyance. The judgments by the subjects were averaged for each sound, resulting in the values shown in the table on the right (high values indicate a high degree of annoyance).

Based on these judgments, we now want to create a metric that allows us to predict the degree of annoyance of other sounds by means of a measurement-based analysis, without having to conduct another jury test. A comprehensive signal analysis

Test signal	Annoyance (average)
M1	7.43
M2	2.29
M3	7.07
M4	3.79
M5	4.93
M6	7.86

Table 1: Results of a jury test

performed in advance using ArtemiS SUITE showed that several psychoacoustic analyses reflect the perceived differences between the sounds quite well. To create the metric, we will use the semi-automatic mode of the Metric Project, which is capable of directly evaluating the results of the SQuare jury test.

Using the Metric Project

A new Metric Project can be opened via **START → New → Metric Project**[2]. A Metric Project consists of several areas (see Figure 1):

- **Sequences**: In this area, you specify the calculation of single-value parameters that could be used for the metric.
- **Metric**: On the right side of the Metric Project, you select the mode used to create your metric. This is also where you specify the desired combination and weighting of the single-value parameters.
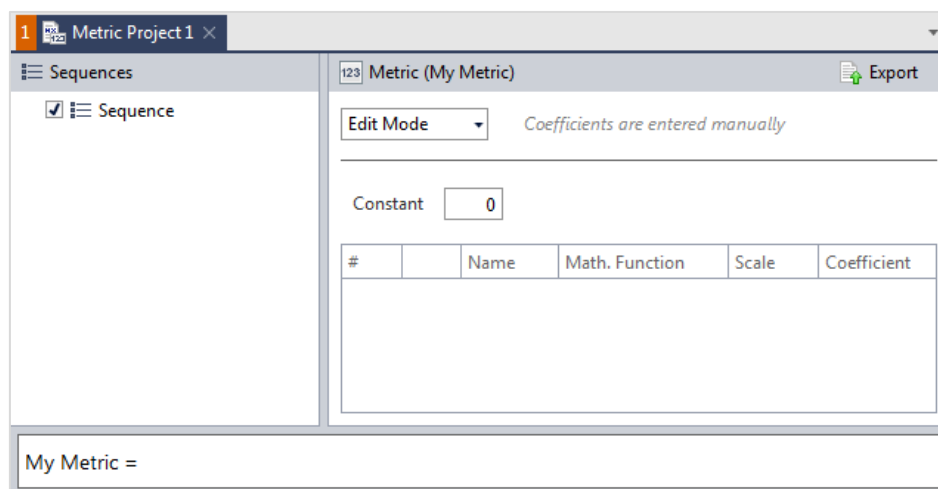- The bottom area displays the mathematical formula of the resulting metric.

Figure 1: Metric Project

---

[2] In order to use a Metric Project, you need a license including **ASM 27** (ArtemiS SUITE Calculation Module, code 5027).

Specifying the properties of the metric

In the properties of the Metric Project (see figure 2), you can specify some basic parameters of your metric. The Properties window can be opened via **View → Properties**).
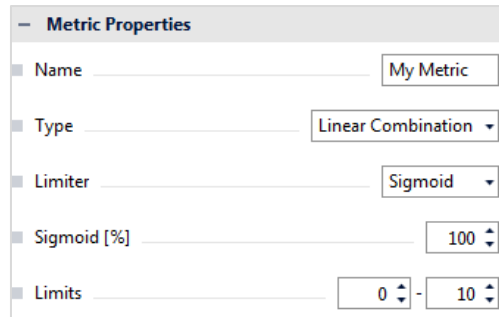
Figure 2: Properties window of the Metric Project

The following parameters of the metric can be specified in this window:
- **Name**: Enter a name for your new metric in this field.
- **Type**: This is where you specify how the individual terms of the metric are combined mathematically:
  - **Linear Combination**: Addition of partial results, e.g., $a + b + c$
  - **Ln (Linear Combination)**: Addition and subsequent taking of the natural logarithm, e.g., $ln(a + b + c)$
  - **Exp (Linear Combination)**: Addition and subsequent use as the exponent of the natural exponential function, e.g., $e^{(a+b+c)}$
  - **Linear Combination²**: Addition and subsequent squaring, e.g., $(a + b + c)^2$

  This selection affects the curve shape of the resulting function. For example, a linear combination results in a less steep rise of the curve than a squared linear combination. This allows you to compress or expand the range of values. In addition, you can specify a mathematical weighting function for each individual term of the metric.
- **Limiter**: This selection allows you to limit the range of values of your metric:
  - **None**: The range of values is not limited.
  - **Hard Limit**: The range of values is limited to fixed borders. All results smaller than the lower limit are set to this limit value, and all results above the upper limit are set to the upper limit value. Results between the limits remain unaffected.
  - **Sigmoid**: This option limits the range of values to certain borders as well. However, a sigmoid function is used for the limitation, which asymptotically approaches the limit values, thus achieving a "soft" (gradual) limitation. You can adjust the strength of this effect with the value in the **Sigmoid [%]** field, where **0** is equivalent to hard limits.

  If you have specified a categorical scale for your jury tests (e.g., the 10-step scale according to the VDI guidelines), a limitation of the range of values is useful to ensure that ratings calculated via your metric fit that scale and do not exceed its range.

Specifying the processing sequences

By means of processing sequences, you can determine which single-value results you want to include in your metric. Please note the following when specifying your processing sequences:
- A metric can only use single-value results from analyses. This means that each sequence must result in a single value at the end. Result curves, such as an **FFT vs. Time** analysis, cannot be used directly for a metric. To create a single-value result, you can either use a single-value analysis or a 2D analysis in connection with the **Single Values** process. You can also use a 3D analysis in connection with the processes **Cut 2D from 3D** and **Single Values**.
- Each processing sequence should only pass the single-value result(s) of one analysis to the metric. If, for example, a sequence contains the processes **Loudness vs. Time** and **Single Values**, you can use it to calculate multiple single-value results for loudness (e.g., average,

median, and maximum), but in order to calculate a sharpness value, you would have to create a separate processing sequence.

- Create separate processing sequences, one for each analysis that might be relevant for the assessment of your sounds. Once calculated, the results are available to you for developing the metric, allowing you to examine their actual influence. However, not all analysis results need necessarily be included in your metric. Only at the end you decide which analyses you will actually use.

Figure 3 shows an example of a Metric Project with four processing sequences. In addition, the Properties window of the **Single Values** element of the processing sequence with the loudness analysis (*Loud*) is shown. In the properties, you can activate multiple single-value results for the calculation. They are then individually available for creating the metric.
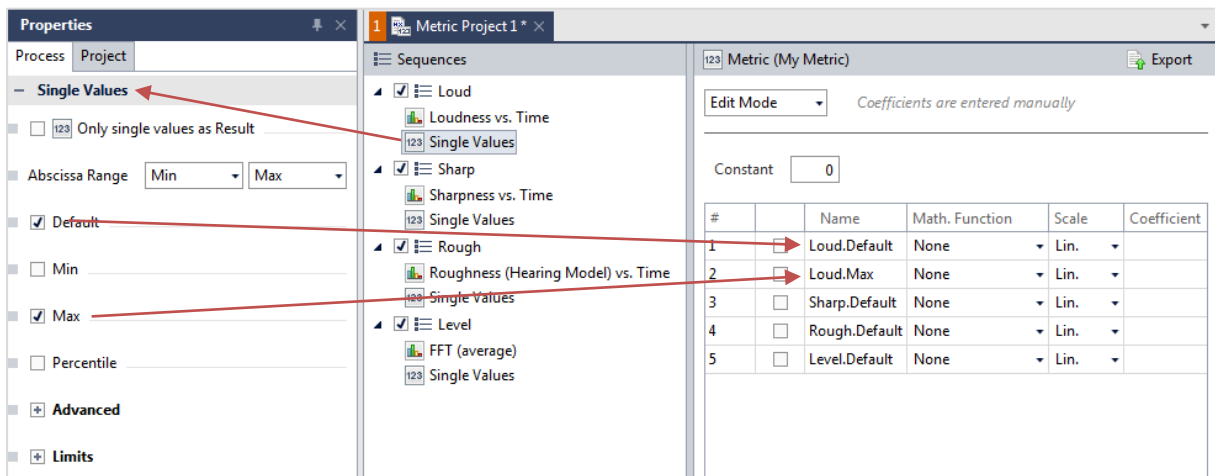

Figure 3: Metric Project and Properties window of the **Single Values** element

Specifying the metric
In a Metric Project, you can create the metric either manually (**Edit Mode**) or semi-automatically (**Design Mode**). Select the mode in the first field in the **Metric** area.

- **Edit Mode** provides you with all single values resulting from your processing sequences. From these values, you build your metric by activating the desired values in the table and specifying a number of mathematical rules (e.g., mathematical functions, such as exponential or logarithmic). This mode is recommended if you already have a calculation rule at hand (e.g., specified by a customer) and you want to model it as a program to be calculated by ArtemiS SUITE.
- In **Design Mode**, you use the results from a jury test to create a metric. ArtemiS SUITE supports you in mapping the jury test results to the measurement-based analysis results in the best possible way.
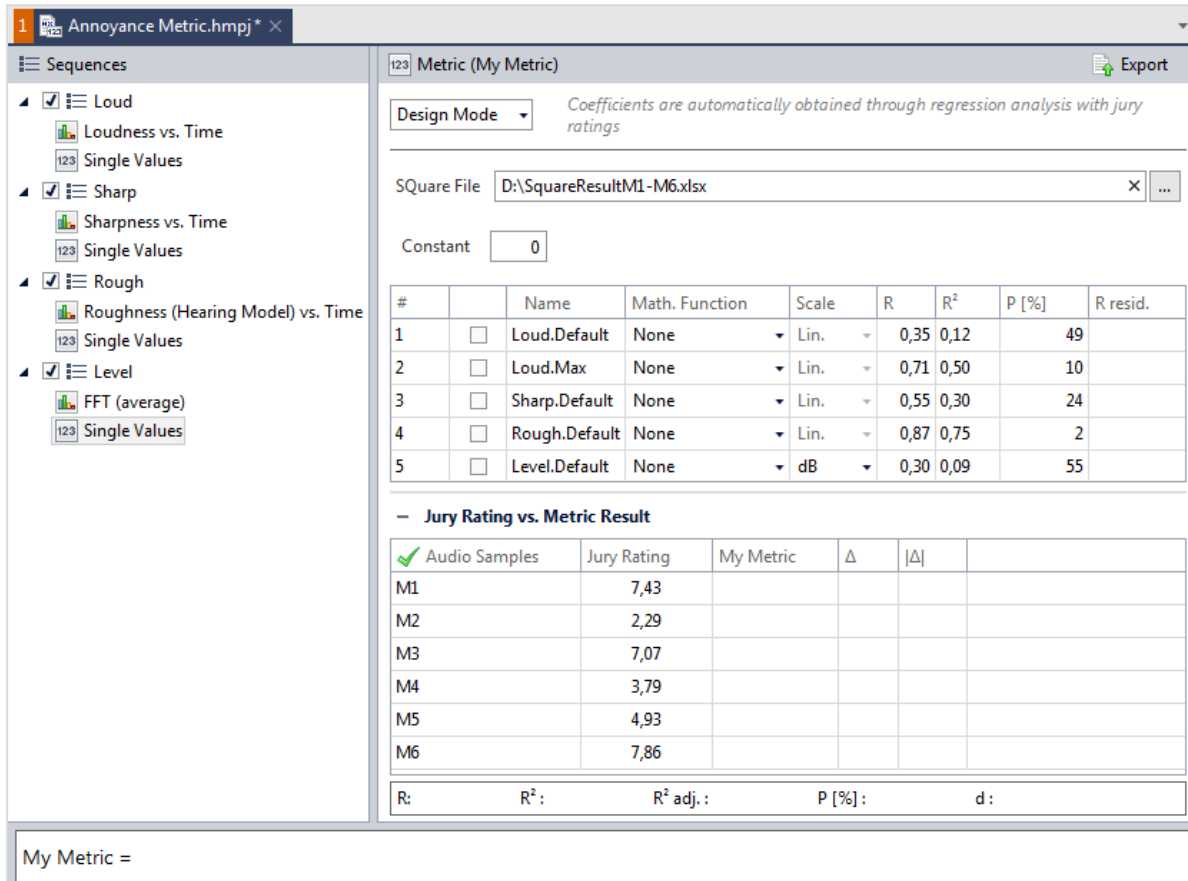
This Application Note describes how to use the Design Mode. More information on creating a metric in Edit Mode can be found in the Help System of ArtemiS SUITE.

Once you have selected Design Mode, the user interface of the Metric Project changes and allows you to load a SQuare file[3] containing the results from your jury test. In order for the Metric Project to work correctly, the HDF files used for the jury test must reside in the same folder as the specified Excel file. You can then have the corresponding entries displayed in a separate table below the table with the list of single-value results (you may need to click on the ⊞ icon). If all referenced HDF files are found, the table displays a ✔ icon. While it is possible to create a metric even if not all HDF files are available, this

---

[3] You can also load an Excel file created otherwise, but it must meet certain requirements (see help system of ArtemiS SUITE → Metric Project).

would lead to some jury test results not being taken into account, which may result in an inadequate metric.

Now click on the ▦ icon in the toolbar of the main window to calculate the single-value results for all analyses and all HDF files (each using the first channel) as well as the corresponding correlation coefficients (see figure 4).

Figure 4:   Metric Project in Design Mode

Besides the single-value results available for creating the metric, the upper table also shows a number of statistical parameters. These parameters are determined automatically and show you to which degree each single-value result correlates with the results of your jury test. The columns of the table are described in more detail in the following:

- **#**: These numbers reflect the order in which the single-value results were added to the list.
- **Active**: Checking the checkbox of a single-value result activates this value for the metric.
- **Name**: This is the automatically generated name of the single-value result, by which you can refer to it in your corresponding processing sequence.
- **Math. Function**: In this column, you can specify how strongly the respective single-value result should affect the final result. Selecting **None** causes the value to be used in the formula without any changes. If you select **Exp**, the value is used as the exponent of the natural exponential function: e^x. With **Ln**, the natural logarithm of the value is taken in the formula: $ln(x)$.
- **Scale**: Your selection in this column determines whether the single-value result is used as a **Linear** value or as a **dB** value in your metric. The **dB** setting is suitable, for example, for single-value results of a level analysis, whereas you should use the linear value in case of psychoacoustic analyses.
- **R**: This column shows the automatically determined correlation coefficient. It describes the correlation between the calculated single-value results and the average results from the jury test.

- **R^2**: This column shows the coefficient of determination R^2.
- **P [%]**: The automatically determined P value is the statistical probability that the apparent correlation between the single-value result and the result from the jury test is caused by pure chance. The P value should be as small as possible, since it estimates the probability of an error.
- **R resid.**: Once you have activated a single-value result, this column shows the automatically determined correlation coefficients for the residues of the other single-value results. The residues are the differences between the single-value results calculated by the (current) metric and the results of the jury test.
  A high value in this column means that the single-value result of this analysis has a high correlation with the residues. This means that this single-value result is likely suitable to compensate the prediction errors of the current version of the metric. Activating this single-value result will therefore likely improve the overall correlation of the resulting metric.
  A small value in this column means that activating this single-value result will probably not improve the metric considerably.
- **Infl. [%]**: If you have activated a single-value result, this column shows you how strongly this value influences the overall result. If the percentage is high, this single-value result is decisive for the metric.
- **Coefficient**: Once you activate a single-value result, ArtemiS SUITE automatically calculates a weighting factor for it, which is displayed in this column. The factor is determined by minimizing the mean square error by means of a regression analysis.

The table **Jury Rating vs. Metric Result** shows the following columns besides the audio samples:
- **Jury Rating**: This column shows the results of the jury test extracted automatically from the SQuare file by ArtemiS SUITE.
- **Metric name** (**My metric** in the example): This column shows the result calculated from the HDF file by the current metric.
- **Δ** and **|Δ|**: These columns show the differences between the two results (**Δ**) as well as the absolute values of these differences (**|Δ|**).

Figure 5 shows an example of the interface after activating two single-value results for the metric calculation. Besides the additional values added to the two tables automatically, the formula of the current metric is displayed.
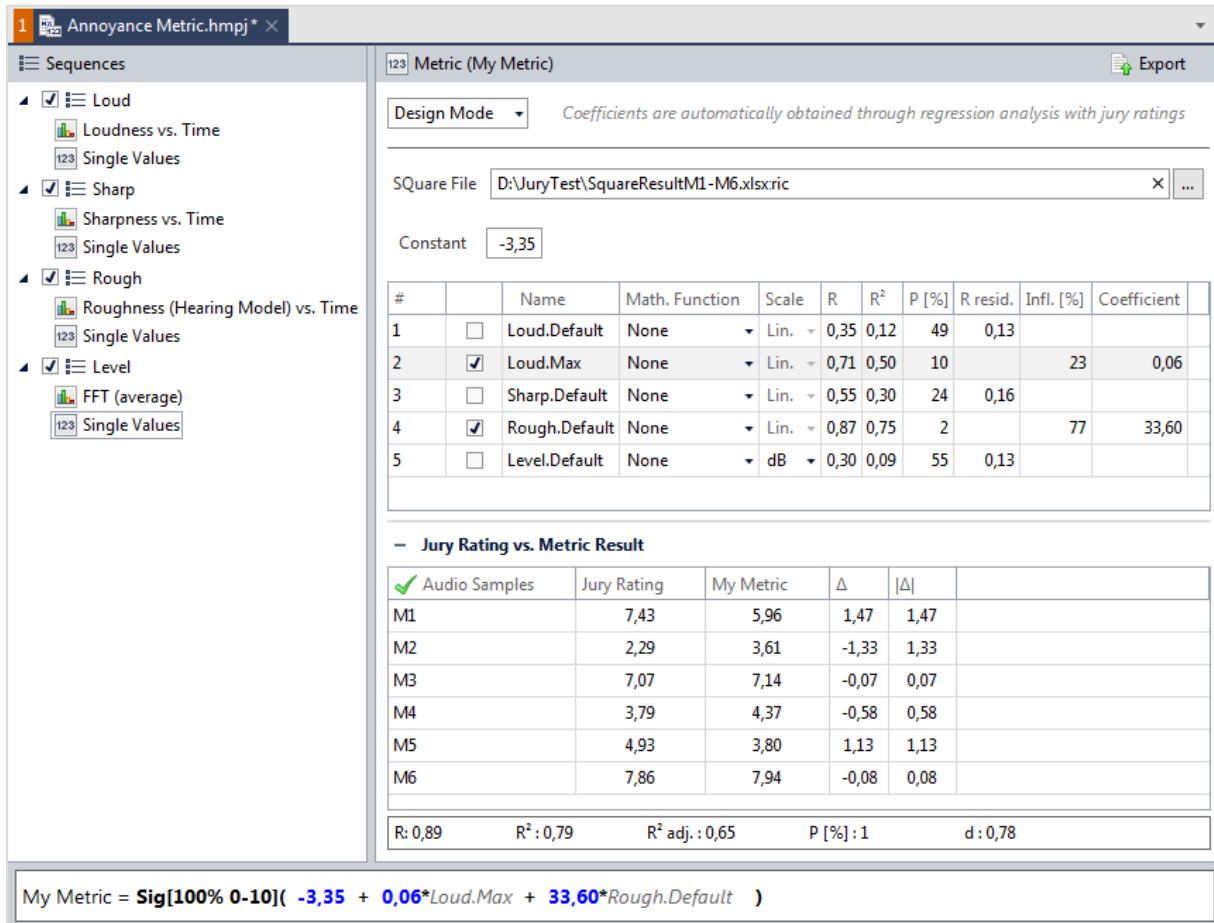
Figure 5:   Metric Project after activating two single-value results

Using the Metric Project, you can now examine different combinations and weightings of single-value results. You can immediately see the effects on the various statistical parameters and on the overall formula. In Figure 5, the lower table shows that the difference between the jury test result and the metric result is very small for some sound samples (M3, M6), whereas it is greater than one category for other samples (M1, M2). This means that the metric does not yet model the results of the jury test very well, and it should be improved by adjusting the parameters or by adding additional single-value results. Figure 6 shows an Excel[4] diagram displaying the jury ratings vs. the metric results derived from figure 5.

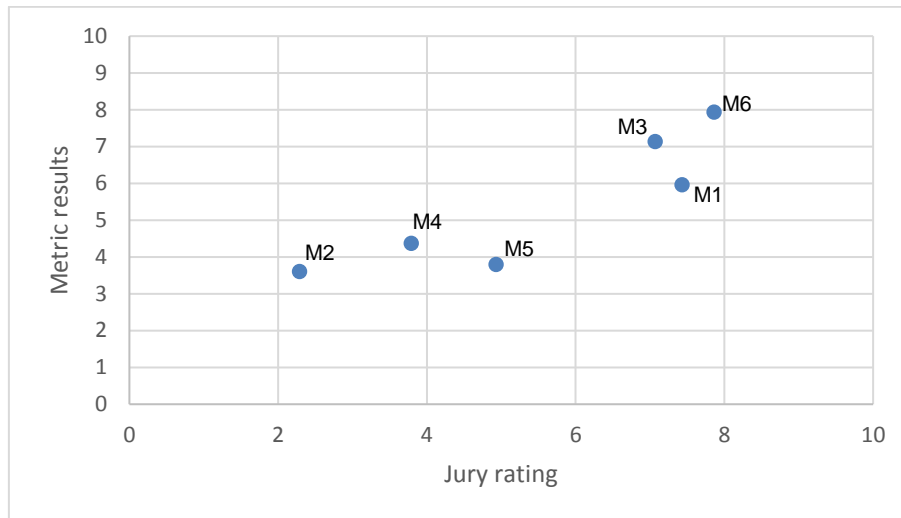[4] Excel is a registered trademark of the Microsoft Corporation.

Figure 6: Metric results vs. jury rating

When you are satisfied with the degree of congruence between the results of the jury test and those calculated by your metric, you can save the Metric Project. Please read the notes in the section "Important notes for creating metrics" at the end of this document. Clicking on the ⬆ button exports your metric specifications to an HMSX file. This file contains all required information about your metric allowing you to apply it to other sound samples. To do so, you can load the metric specification into the Analysis Pool of a Pool Project or into a processing sequence of an Automation Project, a Standardized Test Project or a Metric Project. As an example, the next section describes how to use a metric in a Pool Project.

Using the metric in a Pool Project

In order to use a metric specification in a Pool Project, add a ***Metric*** element to the Analysis Pool (context menu of the Analysis Pools: ***Insert → Single Values → Metric***). In the Properties window of this element, you can select the desired HMSX file after clicking on the ⬚ button. The sound samples to be evaluated by the metric must be added to the Source Pool. Make sure that the sounds have similar characteristics to those you used to create the metric. Finally, you need to add an element to the Destination Pool, which allows the representation of single-value results, such as the ***Single Values Table***. Figure 7 shows an example of a Pool Project with a ***Metric*** element and six new sound samples (M7 to M12).

Clicking on the ▦ button starts the calculation. The resulting single values table for the example used in this Application Note is shown at the bottom of figure 7.
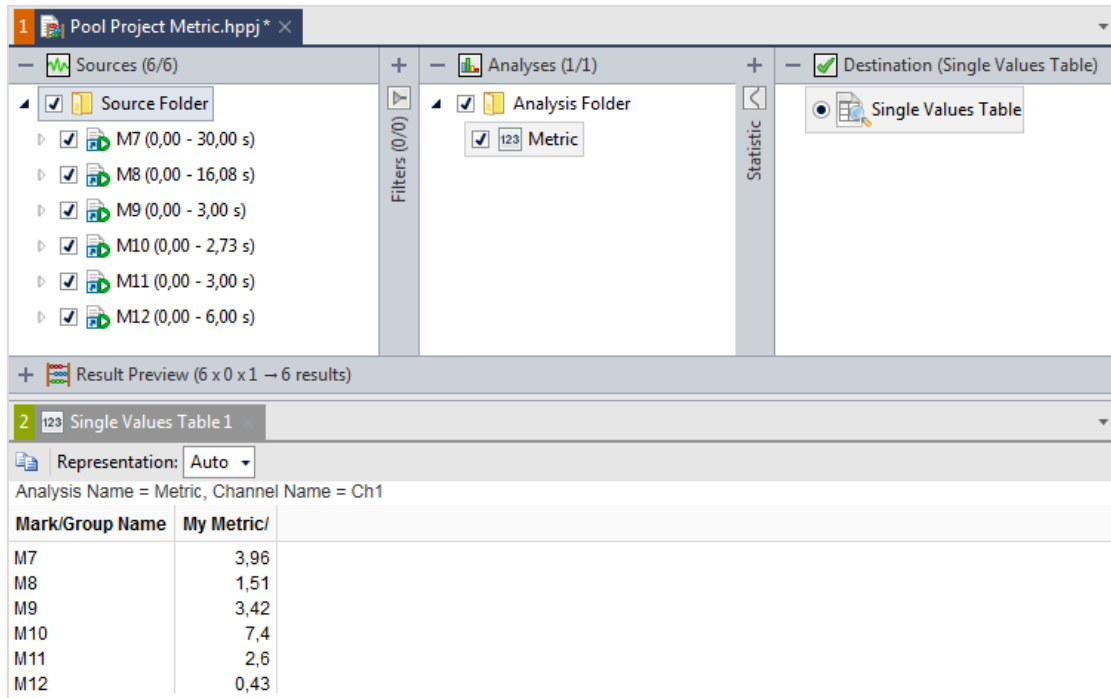
Figure 7: Pool Project with a Metric element in the Analysis Pool

Important notes on creating metrics

In order to obtain a robust metric that is actually suitable to replace jury testing, please assure the following:

- If the metric is based on jury test results, the jury tests must be conducted with a sufficient understanding of sound perception. Only this will ensure that the results are valid and applicable outside the jury test. Otherwise, your resulting metric created from unreliable results of a jury test is based on a bad foundation and will not deliver valid results. Only if the basis, i.e., the results of the jury test, are obtained in a valid way, can the resulting metric deliver meaningful results. Example: You perform a jury test, in which you ask the subjects to rate the sound quality of seat adjustment motors. The sounds used for the jury test were acquired with different recording systems, so that the sounds not only differ from motor to motor, but also due to the recording equipment used. This causes the subjects to include the recording quality in their judgments rather than just the actual motor sound quality. That way, the results of the jury test do not reflect the actual object of examination. A metric based on these jury test results will therefore not deliver valid results.

  In the statistical evaluation of the jury test results, it is important to make sure that significant information is not simply "averaged out". Example: You have performed a jury test, and in spite of a careful description of the task, your subjects have difficulties rating certain sounds. This leads to one group of persons rating these sounds as very good, while the other group rates them as very bad. If you simply averaged these results, these sounds would receive a medium rating. However, such an average rating would not reflect the actual ratings by the subjects. A metric based on these values would not give you a good prediction of the perceived sound quality. In such a case, you must consider which results you want to take into account for your group, and the ratings of the other group must be discarded. You may also have to redesign the test and perform another jury test to verify the results, or determine the reasons for the opposite ratings by the subjects.

  When planning a jury test, you must take into account a number of factors, such as context effects, and carefully select the sound samples and the subjects according to your object of

examination. Planning and performing a jury test costs time and money. Investing these resources is only worthwhile if the jury tests deliver solid results and are actually suitable for creating a metric.

More information on performing jury tests can be found in the sources listed below in the "Literature" section.

- When creating a metric, it is of course important to achieve a high correlation between the results of the jury test and the results of the measurement-based analysis. However, some aspects must be considered here, since a high correlation of results is only a necessary, but not a sufficient, condition. A high correlation should not be considered the only optimization criterion. Instead, the actual goal must be the creation of a valid metric, which delivers meaningful predictions not only for the results of the jury test, but also for other sound samples. In order to avoid an apparent correlation caused by happenstance, the influence of each single-value result must be examined systematically and separately. Each single-value result must be selected regarding its causal influence on sound quality. Example: If you want to examine only sounds mainly consisting of random noise, a high correlation with the single-value results of tonality can only be caused by chance. In spite of its apparently high correlation (spurious correlation), this single-value result should not be included in the metric.

  Likewise, you should avoid using too many different single-value results for your metric. While this might increase the correlation with the examined jury test results, it will not necessarily improve your metric. In [1], you will find a number of statistical parameters you can use to check the robustness of your metric.

  A useful method for creating a robust metric is to subdivide your data into two groups on a random basis. You can then create your metric based on the first data set and test it with the second data set. Since you have perceptive ratings at hand for the second data set as well, you can easily compare the calculated results to the results of the jury test, thus validating your metric. If you not only paid attention to a high correlation when creating your metric, but also based it on plausible, meaningful single-value results, your metric should be able to predict the validation data with a high correlation. In [1], you will find some important notes on this aspect of metric creation.

- The obtained metric can only be used for sounds with similar characteristics to those used for developing the metric. Using the metric for other types of sounds will not lead to meaningful results in many cases. Example: For the jury tests and for creating your metric, you used acceleration sounds of sports cars. The resulting metric can predict the perceived sound quality of comparable sounds very well. However, the metric will fail if you use it to assess the idle engine sound of luxury cars. Even though the sounds are generated by combustion engines in both cases and are measured at comparable positions (e.g., the driver's position in the cabin) and with comparable equipment, these sound types are hardly comparable and cannot meaningfully be examined with the same metric.

Literature

[1]   Andre Fiebig, Fabian Kamp
        "Development of metrics for characterizing product sound quality"
        Proceedings, Aachen Acoustics Colloquium 2015, 123-133
[2]   Imke Hauswirth
        HEAD acoustics Application Note "Conducting jury tests"
        Available at http://head-acoustics.de/eng/nvh_application_notes_jury_evaluation.htm

Do you have questions or suggestions?
Please contact us at imke.hauswirth@head-acoustics.de.
We look forward to your feedback!