Calculating psychoacoustic parameters in ArtemiS SUITE

Psychoacoustics is the science of the relationship between physical quantities of sound and subjective hearing impressions. To examine these relationships, physical parameters, such as sound pressure level, frequency and modulation depth, are mapped to hearing-related parameters. Unlike the physical quantities, these hearing-related quantities – also referred to as psychoacoustic parameters – provide a linear representation of human hearing perception. This means that a doubling of a psychoacoustic quantity corresponds to a doubling of the corresponding subjective perception level.
ArtemiS SUITE offers the possibility to calculate various psychoacoustic parameters. This Application Note explains how the psychoacoustic quantities roughness, fluctuation strength and tonality can be calculated and used in ArtemiS SUITE.

The psychoacoustic parameters loudness and sharpness have already been described in the Application Note “Psychoacoustic Analyses I”, which you can download in the Download Center of our web site.

Calculating Roughness

The roughness parameter is used for the subjective judgment of sound impressions and for sound design. With increasing roughness, noise emissions are perceived as increasingly noticeable and usually as increasingly aggressive and annoying, even if, for example, the loudness or the A-weighted sound pressure level remain unchanged.
The impression of roughness occurs whenever a time-variant envelope exists within a critical band; for example, when tones exhibit a temporal structure due to a variation of their amplitude or frequency. If these variations happen very slowly (below 10 Hz), the human ear is capable of tracking the changes, resulting in an impression of a pulsation or beat. With increasing frequency of the variation, other sound impressions are perceived, such as “R-roughness” (around 20 Hz), which then changes into the actual roughness impression, where the ear is no longer capable of tracking the individual temporal changes. Sounds with envelope variations between 20 and 300 Hz are perceived as rough. Above these frequencies, the main spectral line and sidebands of pure amplitude-modulated tones become audible as individual tones. The roughness depends on the center frequency, the modulation frequency and the modulation depth. The signal level only has a small influence on the roughness impression.
With increasing modulation depth, the impression of roughness becomes stronger. The dependency on the modulation frequency has a band-pass characteristic, i.e. the roughness impression strongly decreases towards very low or high modulation frequencies. In a judgment of various amplitude-modulated sine tones, each having a center frequency of 1 kHz and a modulation depth of 1 (100%), but being modulated at different frequencies, the maximum roughness is perceived at a modulation frequency of about 70 Hz.
For lower carrier frequencies, the maximum shifts towards lower modulation frequencies.
Roughness is not only caused by amplitude-modulated tones, but also by frequency modulation and by amplitude-modulated noise. The unit of roughness is *asper*. A sine tone of 1 kHz with a level of 60 dB, amplitude-modulated at a frequency of 70 Hz and with a modulation depth of 1, is defined to have a roughness of 1 asper.

Basically, a roughness impression can also be caused by two tonal components occurring within a critical bandwidth of human hearing. In communications engineering, this is referred to as “carrier-less amplitude modulation”.

ArtemiS SUITE provides an algorithm, which calculates the roughness based on the hearing model according to Sottek [1]. The analysis ***Roughness (Hearing Model) vs. Time*** simulates the signal processing of human hearing and judges the roughness of a signal in a similar way as the human hearing system. The block diagram shown in figure 1 illustrates the roughness calculation based on the hearing model.
First a filtering of the audio signal takes place in order to account for the influence of the outer and middle ear. Afterwards, the signal is subdivided by a filter bank with parallel, overlapping band-pass filters. The distance between the center frequencies of adjacent filters is constant on the tonality scale. The number of band-pass filters can be set to 24 or 47 filters in the Properties window of the analysis. Using 47 band-pass filters allows a more accurate simulation of the natural hearing process; however, it also increases the calculation time. After filtering, the envelopes of the partial band signals are determined using the Hilbert transformation. To take the threshold in quiet into account, the excitation levels are reduced (approx. 20 dB/decade for frequencies below 500 Hz). In the next processing step, filtering with 3rd order low-pass filters takes place. The cutoff frequency of these filters is frequency-dependent and is about 120 Hz at 1 kHz. The low-pass filtering accounts for the fact that the human ear cannot track the variation of the envelope above a certain rate. Afterwards, the envelope variations are distorted in a nonlinear way. The nonlinear curve used for this is an exponential function with an exponent of 0.125. The next step is calculating the autocorrelation function. Afterwards, the partial roughnesses can be determined by filtering with 3rd-order high-pass filters and an amplification g_R (z_i). Both the cutoff frequency of the high-pass filters and the weighting depend on the frequency position of the analyzed partial band (at 1 kHz, the cutoff frequency is approx. 120 Hz).
The high-pass filtering is necessary to account for the decrease of perceived roughness towards lower modulation frequencies. The combination of the high-pass and low-pass filters models the typical band-pass characteristic regarding the relationship between roughness and modulation frequency. The weighting g_R (z_i) accounts for the influence of the frequency position of the carrier frequency for the roughness impression. In addition, a frequency-band-spanning weighting is applied as described in [2]. A stronger weighting is applied to the roughness of the frequency bands that exhibit a strong roughness. This represents a kind of masking effect: If a particularly high roughness is present in a certain frequency band, this roughness can cover up the roughnesses of neighboring frequency bands; therefore, it must contribute more to the total roughness than the roughnesses of the adjacent frequency bands.
After the weighting, the total roughness is ready to be calculated by integrating the partial roughnesses. A detailed description of the hearing model and the roughness calculation based on it can be found in [1], [2] and [3].

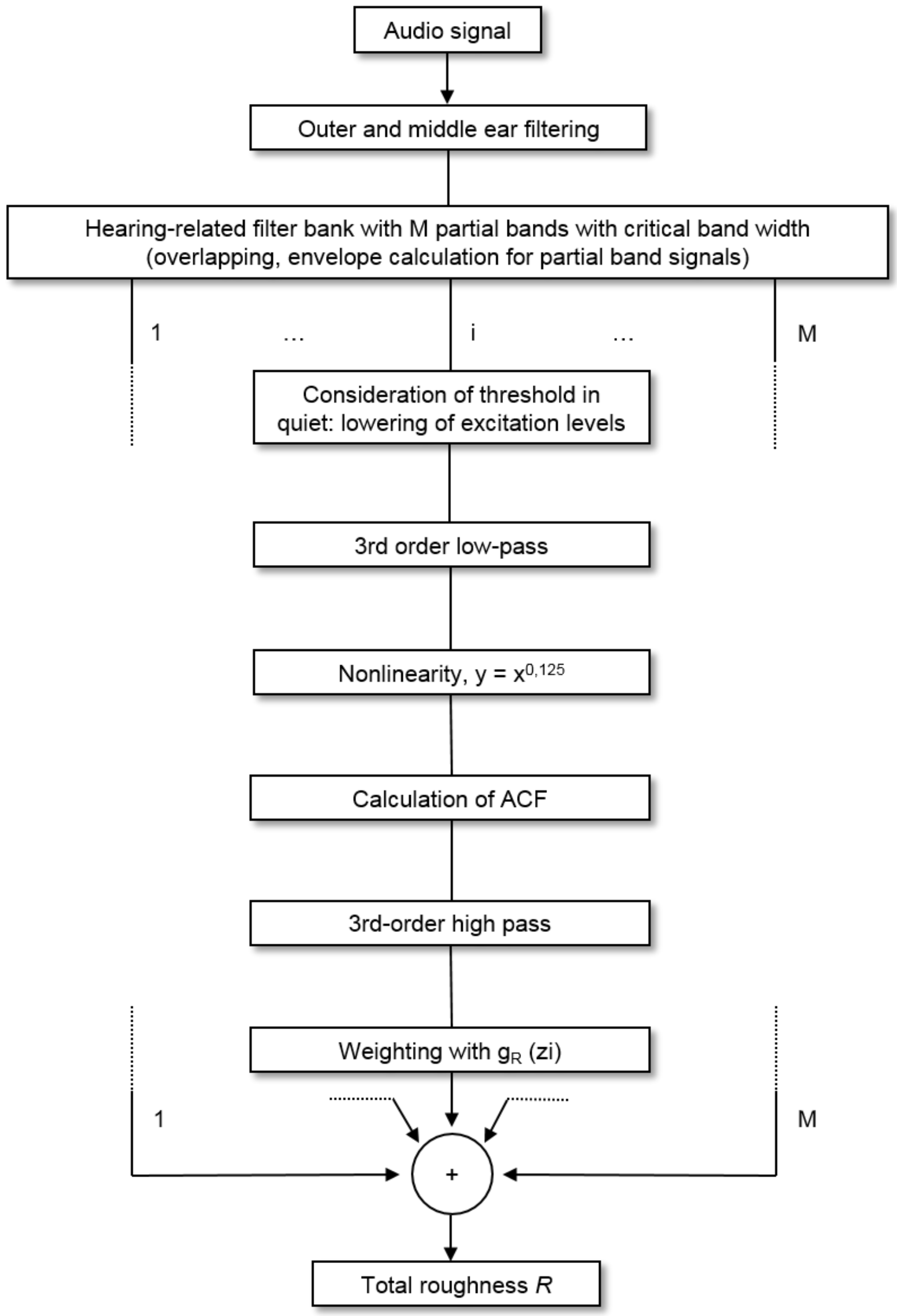

Figure 1: Block diagram of the roughness calculation based on the hearing model according to Sottek (from [3])

Besides normal roughness, literature sources (e.g. [4]) also describe the R-roughness (also referred to as α-roughness or slow roughness). R-roughness occurs at lower modulation frequencies between 15 and 45 Hz and has a maximum at a modulation frequency of approx. 20 Hz.

R-roughness is particularly often perceived in the noise of combustion engines. This kind of noise is a consequence of the occurrence of half engine orders. If these engine orders occur prominently within a critical bandwidth, the engine sound is perceived as "rough". Due to the fact that the combustion processes in the individual cylinders of a combustion engine are not exactly identical (for example, due to variations in the intake system or the exhaust manifold), half engine orders occur in addition to the main or integer engine orders. This means that, for example, a four-cylinder engine will exhibit not only the 4^{th} engine order, but also the 3.5^{th} and/or the 4.5^{th} engine order at certain RPM values. For example, at 3000 rpm, the 4th engine order is 200 Hz, the 3.5^{th} order is 175 Hz and the 4.5th order is 225 Hz.

Depending on the amplitude proportion of the half engine order, the modulation depth will vary and have a corresponding effect on the sound impression.
With the analysis ***Roughness (Hearing Model) vs. Time***, the R-roughness is accounted for insofar as this analysis includes the shift of the maximum of the roughness perception towards slower modulation frequencies at lower carrier frequencies in the calculation. In addition, the R-roughness can also be detected using a modulation analysis in ArtemiS SUITE. This analysis determines the modulation depth depending on the carrier frequency and the modulation frequencies. Regarding modulation analysis, a separate Application Note is available, which you will find in the Download Center of our website.

Application and Examples

Figure 2 shows the Properties window of the analysis ***Roughness (Hearing Model) vs. Time***. For this analysis, two settings can be configured. As explained above, the ***Resolution*** can be specified, i.e. the number of band-pass filters for the hearing model. With the setting ***1/1 Bark***, 24 filters are used. With ***1/2 Bark***, the number increases to 47 filters. Furthermore, the option ***Skip Analysis Start*** can be used to suppress the transient effect of the digital filters at the beginning of the analysis for the specified number of seconds in order to prevent it from distorting the total result. Enabling this function is particularly advisable if a single value calculation is to be performed.

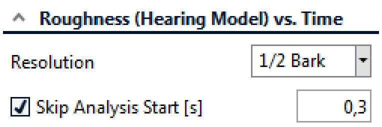

Figure 2: Properties window of the analysis ***Roughness (Hearing Model) vs. Time***

For the diagram in figure 3, the analysis ***Roughness (Hearing Model) vs. Time*** was applied to three synthetically generated signals. These signals are 1 kHz tones modulated with a sine curve at different modulation frequencies and depths.

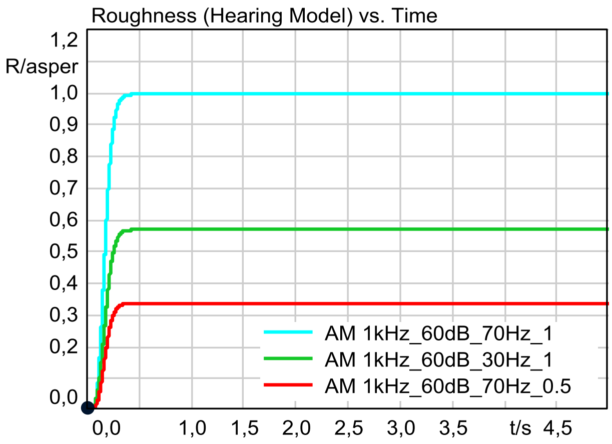

Figure 3: Results of the analysis ***Roughness (Hearing Model) vs. Time*** for synthetic signals

The cyan curve is the result for a 1 kHz tone modulated with a sine-shaped modulation at a frequency of f_m=70 Hz and a modulation depth of m=1 and with a level of 60 dB. According to the definition of roughness, this signal has a roughness value of 1 asper. After a brief settling phase, the curve shows this value. The other two signals differ from the reference signal by a lower modulation frequency in one case (f_m=30 Hz, green curve) and a lower modulation depth in the other case (m=0.5, red curve). The values shown after the settling phase match those stated in [5] fairly well in both results.

Compared to that, actual technical noise signals hardly ever show such high roughness values. Figure 3 shows the results of a roughness analysis of two small electric motors.

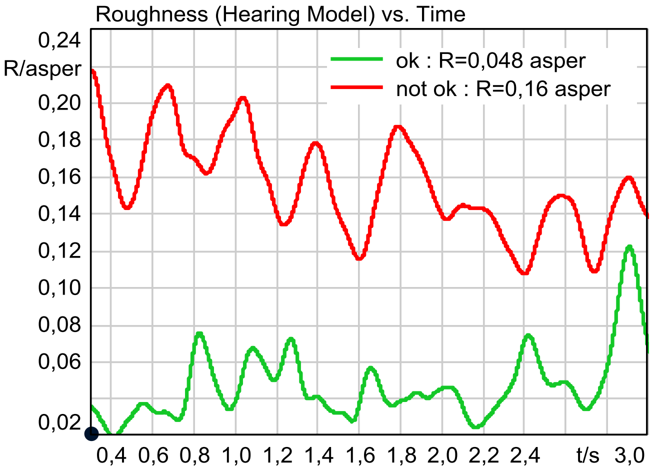

Figure 4: Results of the analysis ***Roughness (Hearing Model) vs. Time*** for actual technical signals

The sounds of both electric motors contain a disturbing component. This component is much stronger in one of the two sounds (red curve), so that it is perceived as very rough and would not pass an acoustic quality check. The other sound has a much weaker disturbing component (green curve). The results show that the analysis ***Roughness (Hearing Model) vs. Time*** delivers smaller values when applied to actual technical signals. However, the difference between the two curves and the resulting single values reflect the hearing impression very well. The single value shown for the worse signal is about three times higher than the value for the better signal. In order not to distort the single value, the first 0.3 seconds of the analysis were suppressed using the ***Skip Analysis Start*** option.

Basically, in a roughness analysis, the absolute values of the analysis results should be considered less important than the ratio between different results.

There are several publications where roughness algorithms are compared and applied to various types of sounds. It turns out that the results of the roughness analysis based on the hearing model shows a good correlation with the subjective judgments obtained in listening tests (see [2], [6], [7]).

Calculating Fluctuation Strength

The impression called fluctuation strength is caused by signal variations with very low modulation frequencies. The maximum of this psychoacoustic quantity is at modulation frequencies around 4 Hz. Just like roughness, the fluctuation strength impression shows little dependency on the signal level. The unit *vacil* is defined by the same sine tone as in the case of roughness, except that the modulation frequency is 4 Hz instead of 70 Hz. The calculation of the fluctuation strength in ArtemiS SUITE is done similarly to the calculation of ***Roughness (Hearing Model) vs. Time***. The algorithm for the calculation of the roughness has been adapted in a way that the maximum of the fluctuation strength is obtained at 4 Hz instead of 70 Hz as for the roughness.

Application and Examples

Just like the analysis ***Roughness (Hearing Model) vs. Time***, the Properties window of the analysis ***Fluctuation Strength vs. Time*** provides a ***Resolution*** field, where the number of band-pass filters for the hearing model can be selected. Furthermore, the ***Skip Analysis*** *Start* option for suppressing the transient effect for the specified number of seconds is available as well (see figure 5).

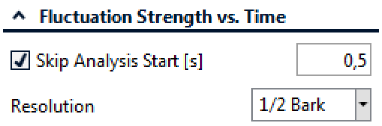

Figure 5: Properties window of the analysis ***Fluctuation Strength vs. Time***

In the following example, the fluctuation strength of two voice signals is calculated. Fluent speech contains about four syllables per second [5]. This means that speech is modulated with a frequency of about 4 Hz and thus has a high fluctuation strength. In order for speech to be well understood, a sufficiently high modulation depth is necessary among other factors. In a loud environment, such as a moving car, the interfering noise reduces the modulation depth of perceived speech, making it harder to understand.
Figure 6 shows the result of the fluctuation strength analysis of the two voice signals. The green curve shows the fluctuation strength of three short sentences spoken in a quiet environment. The red curve shows the analysis results of the same sentences spoken in a moving car.

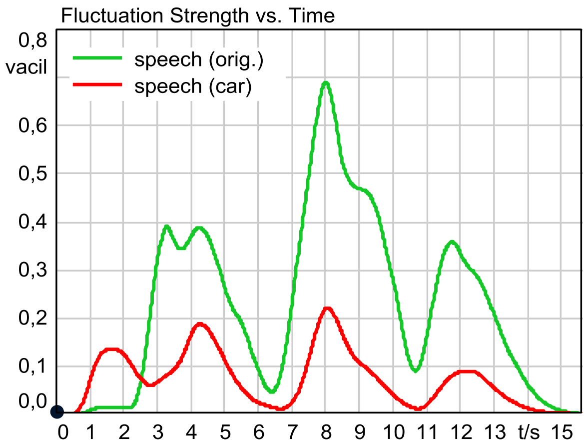

Figure 6: Fluctuation strength of two voice signals (green curve: voice without background noise, red curve: voice in a moving car)

In both curves, the fluctuation strength increases during the sentences and decreases in the pauses. However, the fluctuation strength of the voice recorded in the car has considerably lower values. The fluctuation strength of the voice is reduced, because the background noise reduces the variation of the envelope of the signal and thus the modulation depth.

This example shows that the interpretation of psychoacoustic analysis results strongly depends on the type of sound to be judged. Modulated, i.e. variable sounds command a higher attention than unmodulated sounds. If the listener is interested in the information conveyed in the sound, modulated sounds are not perceived as annoying. On the other hand, if the sound is undesirable, modulated signals are perceived as more annoying than unmodulated signals at the same volume or loudness. For speech, a high fluctuation strength is an advantage, whereas the slowly-modulated noise of, for example, a wind power plant is often judged as annoying.

Calculating Tonality according to DIN 45681

The analysis ***Tonality DIN 45681*** is used to determine the tonal components of noise and to determine a tone adjustment for the assessment of noise immissions according to DIN 45681. This analysis is not a psychoacoustic analysis in the classical sense, because it does not deliver results reflecting the human perception linearly. Nevertheless this analysis will be introduced in the following for the sake of completeness.
The standard describes a method (procedure) for the automatic identification of tones and tone groups from narrow-band spectra. The analysis determines the difference between the tone level and the level of the background noise in the surrounding frequency group, additionally accounting for a frequency-dependent masking effect. From the maximum difference, a penalty value between 0 and 6 dB (in steps of 1 dB) is determined and assigned in order to account for the tonal component(s).

Level Difference	Penalty
> 12 dB	6 dB
> 9 dB	5 dB
> 6 dB	4 dB
> 4 dB	3 dB
> 2 dB	2 dB
> 1 dB	1 dB
<= 1 dB	0 dB

The diagram of this analysis can show the following results:

- Individual tones and their level differences as individual peaks in the spectrum
- Tone groups and their level differences as rectangles in the spectrum
- Tonal penalty as a single value in the header or in the legend
- In the analyses ***Tonality DIN 45681 vs. Time/RPM***, the mean level difference is shown in addition as a single value. It is averaged energetically from the maximum level differences of the individual spectra.

Application and examples

The level differences are determined based on an FFT spectrum. The ***Spectrum Size*** and ***Overlap*** parameters for calculating this spectrum can be configured in the corresponding fields in the Properties window of the analysis (see figure 7).

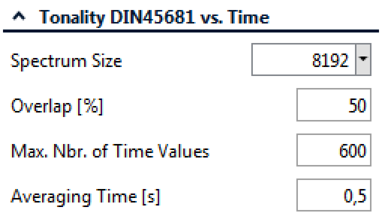

Figure 7: Properties window of the analysis ***Tonality DIN 45681 vs. Time***

According to DIN 45681, the frequency resolution (sampling rate/DFT length) should not exceed 4 % of the frequency group width. That is to say that with a frequency resolution of

$$\frac{\text{samplingrate}}{\text{DFT length}} = \frac{44100\text{Hz}}{8192} = 5.38\text{Hz}$$

only the values above a frequency group width of $\frac{5{,}38\,\text{Hz}}{0{,}04} = 134{,}5\,\text{Hz}$ (thus above 6 Bark) comply with the recommendation of the DIN 45681.

Furthermore, the analysis ***Tonality DIN 45681 vs. Time*** or ***RPM*** supports moving averaging, which can be activated in the Properties window. As soon as a number greater than zero is entered in the ***Averaging Time*** field, the result is averaged over the specified length in order to suppress uncorrelated signal components.
Figure 8 shows an example for the analysis ***Tonality DIN 45681 vs. Time***.

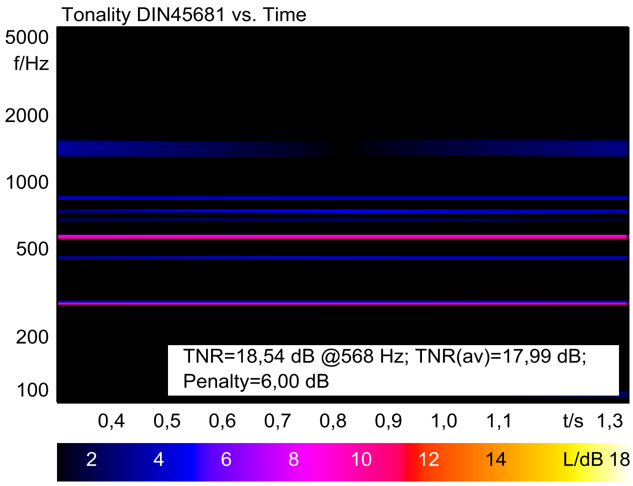

Figure 8: Analysis ***Tonality DIN 45681 vs. Time***

The analyzed sound file contains several tonal components, the strongest of which is the one at 568 Hz, which stands out against the surrounding spectrum by almost 18 dB. If a tonality penalty value was to be assigned to this sound, the maximum value of 6 dB would be used.

Notes

For calculating the analyses presented in this Application Note by means of a Pool Project, you need the following ArtemiS SUITE modules: **ASM 00** ArtemiS SUITE Basic Framework (code 5000) and **ASM 01** ArtemiS SUITE Basic Analysis Module (code 5001). Furthermore, for calculating roughness, you need **ASM 16** ArtemiS SUITE Advanced Psychoacoustics Module (code 5016), for fluctuation strength: **ASM 12** ArtemiS SUITE Psychoacoustics Module (code 5012), and for tonality: **ASM 17** ArtemiS SUITE Advanced Analysis Module (code 5017). If you want to calculate the analyses by means of an Automation Project or a Standardized Test Project, you may need other modules. Your HEAD acoustics representative will gladly provide you with further information.

Literature

[1] R. Sottek
Modelle zur Signalverarbeitung im menschlichen Gehör (in German)
Dissertation, RWTH Aachen, 1993

[2] R. Sottek, P. Vranken, H.-J. Kaiser
Anwendung der gehörgerechten Rauhigkeitsberechnung (in German)
Proc. DAGA ’94, Dresden 1994, 1201-1204

[3] R. Sottek
Gehörgerechte Rauhigkeitsberechnung (in German)
Proc. DAGA ’94, Dresden 1994, 1197-1200

[4] N. Springer, R. Weber
Existenzbereich der R-Rauhigkeit (in German)
Proc. DAGA ’94, Dresden 1994, 1221-1224

[5] H. Fastl, E. Zwicker
Psychoacoustics, Facts and Models, Third Edition
Springer-Verlag, Berlin, Heidelberg, 2007

[6] F. Attia, A. Okker
Unterscheidung zwischen Rauhigkeitsmodellen und Übereinstimmung mit dem subjektiven Gehöreindruck (in German)
Proc. DAGA ’95, Saarbrücken 1995, 843-846

[7] M. Klemenz
Die Geräuschqualität bei der Anfahrt elektrischer Schienenfahrzeuge (in German)
Dissertation, RWTH Aachen, 2005

Do you have any questions or comments?
Please write to imke.hauswirth@head-acoustics.de.
We look forward to receiving your feedback!