Psychoacoustic Analyses in ArtemiS II

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 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. The psychoacoustic parameters loudness and sharpness have already been described in the Application Note "Psychoacoustic Analysis I", which you can download, in the Download area of our web site.

Calculating Roughness

Introduction

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 provides two roughness analysis functions based on different approaches. The algorithms used are explained in the following section.

The “Roughness vs. Time” analysis calculates partial roughnesses from the modulation depths of partial signal bands and adds them up to determine the total roughness. The signal is first subdivided into 24 partial bands by a linear-phase filter bank, each band being 2 Bark wide and overlapping by 1 Bark. This larger bandwidth – compared to the loudness calculation – is necessary to detect sufficiently high modulation frequencies that are still within one band. The overlapping assures that sound events at the edges of the bands do not lead to completely different results than those nearer the center of the bands.

A partial roughness value is calculated from each partial band signal. This is achieved by calculating the envelope of the partial band signal using a Hilbert transformation. The result is filtered with IIR filters modeling the dependency of the roughness on the modulation frequency.

Then the modulation depth values m_i are calculated with

$$m_i = \sqrt{2\frac{P_{\sim,i}}{P_{-,i}}}$$

where $P_{-,i}$ is the power of the constant component of the partial band signal before the IIR filtering and $P_{\sim,i}$ is the power after the filtering. The Integration time for determining the power values is 100 ms.

The partial roughness is calculated as

$$r_i' = \begin{cases} K(m_i k_i)^{1.5} & for\ m_i < 1 \\ K m_i k_i & for\ m_i \geq 1 \end{cases}$$

Here, k_i is a factor representing the dependency of the roughness on the frequency position of the partial band (see [1], page 61). K is a scaling constant reflecting the fact that 1 asper is equivalent to a 1 kHz sine tone with a level of 60 dB amplitude-modulated at a rate of 70 Hz and with a modulation depth of 1.

The dependency of the roughness on the sound pressure level is taken into account by doubling the partial roughness for each 20 dB increase of the constant power component of the partial band signal. The total roughness is the sum of all partial roughnesses.

One problem with this method of roughness calculation is that the analysis of signals with unmodulated noise yields roughness values that are much too high compared to the actual perception. In order to avoid this problem, new approaches for the calculation were investigated.

In addition to the “Roughness vs. Time” analysis, ArtemiS provides another algorithm, which calculates the roughness based on the hearing model according to Sottek [2]. The analysis “Hearing Model Roughness 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.

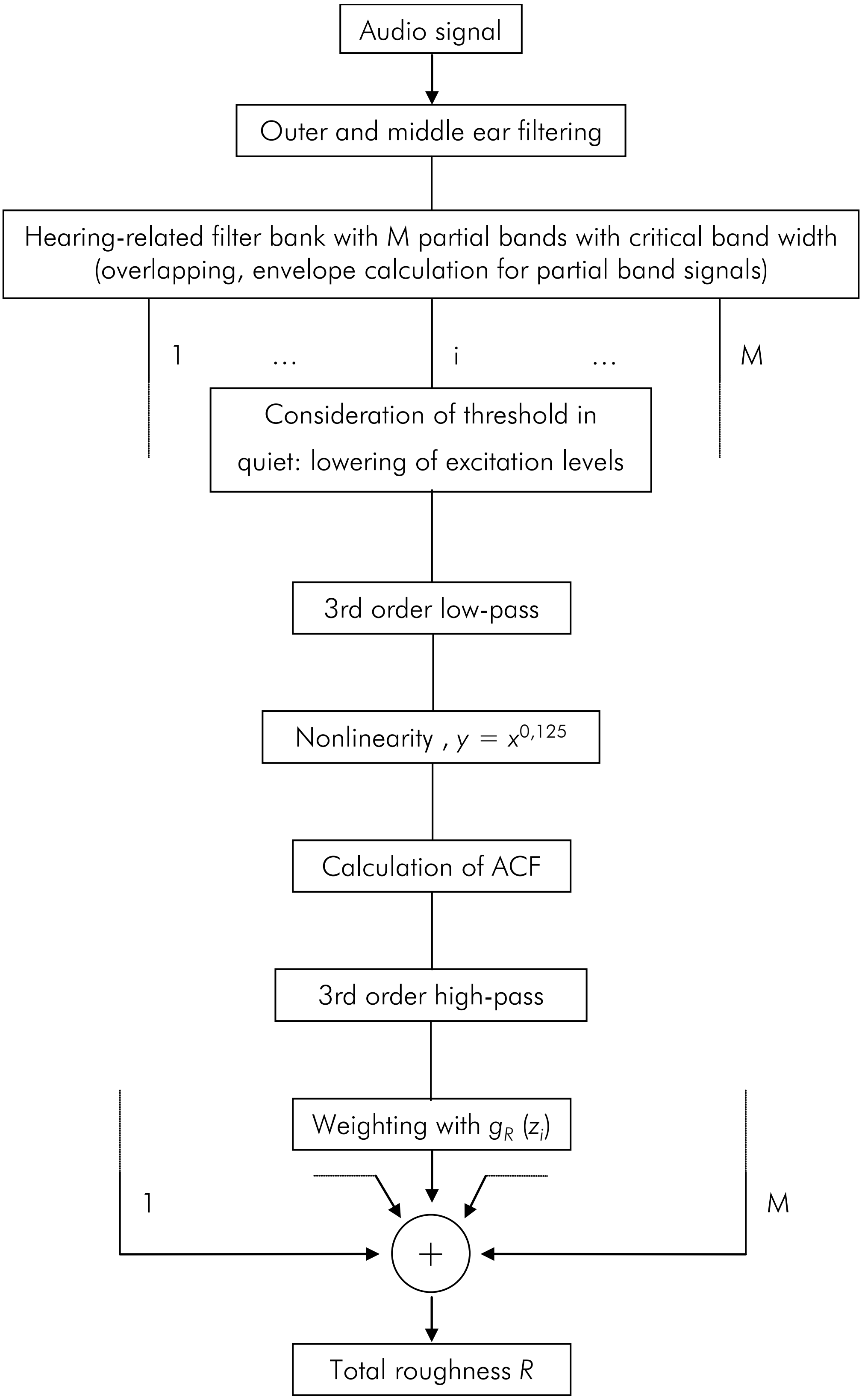

Figure 1: Block diagram of the roughness calculation based on the hearing model according to Sottek (from [3])

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 selected in the Properties dialog of the analysis under “Resolution”. With the setting “1/1 Bark”, 24 filters are used. With the setting “1/2 Bark”, the number is increased to 47 filters. The higher number of 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 [4]. 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 rough-nesses.

A detailed description of the hearing model and the roughness calculation based on it can be found in [2], [3] and [4].

Besides normal roughness, literature sources (e.g. [5]) 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 4th engine order, but also the 3.5th and/or the 4.5th engine order at certain RPM values. For example, at 3000 rpm, the 4th engine order is 200 Hz, the 3.5th 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.

Since the ArtemiS analysis "Roughness vs. Time" is standardized in a way that the roughness at a modulation frequency of 70 Hz has the strongest weighting, the R-roughness cannot be represented correctly with this analysis. With the analysis "Hearing Model Roughness 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. 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 area of our website.

Application and Examples

In the following section, we will first apply the analysis functions "Roughness vs. Time" and "Hearing Model Roughness vs. Time" to three synthetically generated signals. These signals are 1 kHz tones modulated with a sine curve at different modulation frequencies and depths. In figure 2, the left diagram shows the results of the "Roughness vs. Time" analysis, the right one shows the results of the "Hearing Model Roughness vs. Time" analysis.

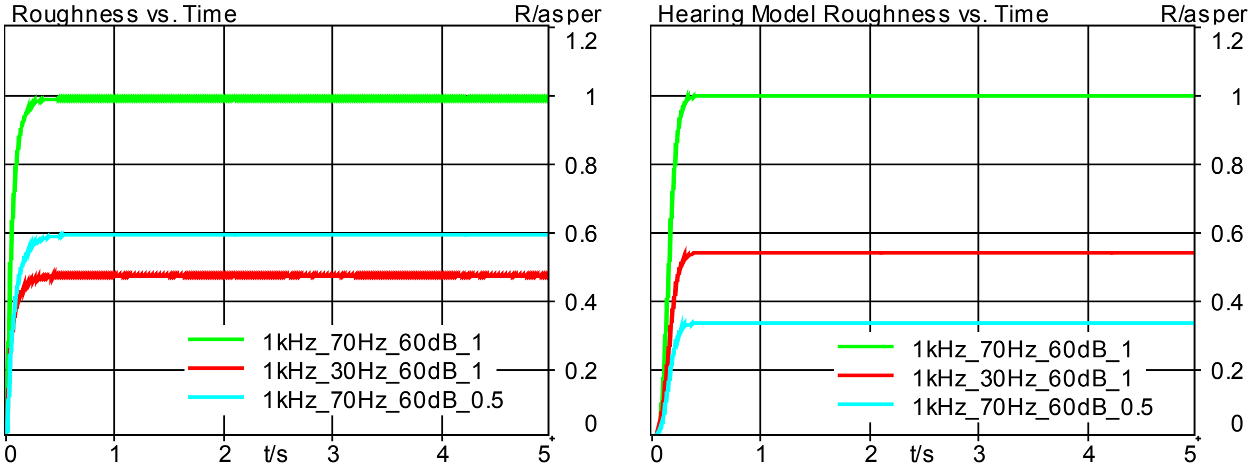

Figure 2: Comparison of the analysis results for synthetic signals (left: Roughness vs. Time; right: Hearing Model Roughness vs. Time)

The green curves are the results for a 1 kHz tone modulated with a sine-shaped modulation at a frequency of 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, both results show the same value. The other two signals differ from the reference signal by a lower modulation frequency f_m=30 Hz in one case (red curve) and a lower modulation depth m=0.5 in the other case (cyan curve). The values shown after the settling phase match those stated in [6] fairly well in both results – but the results of the "Hearing Model Roughness vs. Time" analysis are closer to the results stated in literature and also have the correct order.

A comparison of the analysis results for real-world sounds shows more significant differences. Figure 3 shows the roughness analysis results for two small electric motors. Again, the left diagram shows the results of the "Roughness vs. Time" analysis, the right one shows the results of the "Hearing Model Roughness vs. Time" analysis.

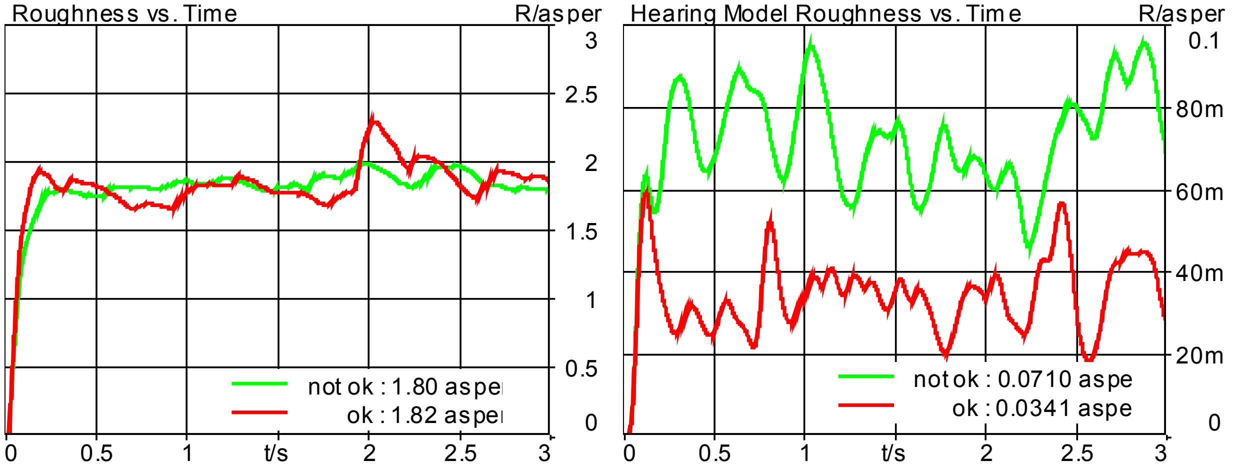

Figure 3: Comparison of the analysis results for real world signals (left: Roughness vs. Time; right: Hearing Model Roughness vs. Time)

The sounds of both electric motors contain a disturbing component. This component is much stronger in one of the two sounds (green 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 (red curve). A comparison of the analysis results shows the following: Even though the asper values of the “Hearing Model Roughness vs. Time” analysis are much lower, this analysis clearly reflects the actual hearing impressions. In the hearing model analysis, the displayed single value of the worse signal is twice as high as that of the better signal. By contrast, the single values of the roughness analysis in the left diagram hardly differ. Furthermore, these values are much too high compared to the 1-asper reference signal. In reality, such high roughness values hardly occur with technical devices and actual sound sources.

The above example shows that the “Hearing Model Roughness vs. Time” analysis provides a better representation of the perceived roughness of real-world sounds. In a roughness analysis, the relative ratio of the single values to be compared should be given a higher importance than the absolute values.

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 [4], [7], [8]).

Calculating Fluctuation Strength

Introduction

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 is done similarly to the calculation of “Hearing Model Roughness 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

In the following example, the fluctuation strength of two voice signals is calculated. Fluent speech contains about four syllables per second [6]. 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 4 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.

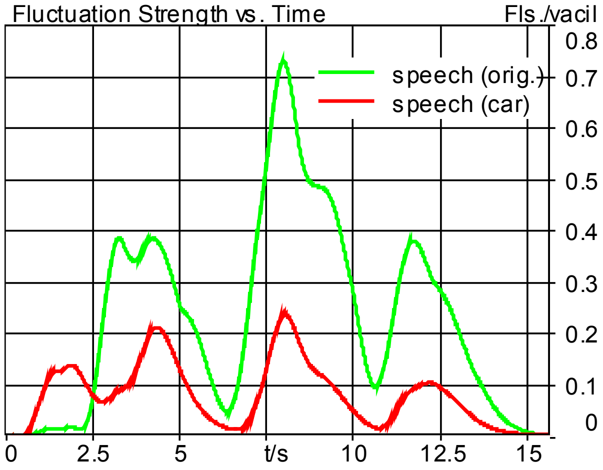

Figure 4: Fluctuation strength of two voice signals (green curve: voice without background noise, red curve: voice in a moving car)

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

Introduction

The tonality of a sound indicates whether the sound consists mainly of tonal components or broadband noise. The contribution of tones to tonality depends on their frequency; at about 700 Hz, the maximum tonality impression is reached. Narrow-band noise with a bandwidth smaller than 1 Bark is also perceived as tonal, though to a decreasing degree with increasing bandwidth. The unit of tonality, *tu* (tonality unit), is defined for a 1 kHz sine tone with a level of 60 dB. The calculation of tonality in ArtemiS is based on publications by E. Terhardt and

W. Aures ([1], [9]). Terhardt uses the algorithm to determine the pitch of sounds, whereas Aures uses it to calculate a psychoacoustic quantity that contributes to the euphony of sound events.
The calculation of tonality is based on short-term spectra obtained with an FFT over 4096 sampling points and a Hanning window. In the first step, the algorithm looks for spectral lines S_i that are higher than their two neighbors $S_{i\pm1}$. Of these lines, only those that are at least 7 dB higher than the lines $S_{i\pm2}$ and $S_{i\pm3}$ are taken into account. The groups of seven lines (with the indices i-3 to i+3) that are found this way are treated as pure tones and removed from the spectrum. For each critical band, a maximum of one tonal component is accounted for. In the remaining spectrum, the algorithm then searches for narrow-band noise[1] with a bandwidth smaller than the critical bandwidth at this position. This is done because such signals also cause an impression of tonality, though to a lesser degree than pure tones.
For the identified tonal components, the sound level excess $L_{\Delta,i}$ is calculated. It is the level of the component minus

- the threshold in quiet,
- the noise power in the respective critical band, defined as the power of the remaining spectral lines after the removal of the tonal component,
- and the excitation level, which results from the other tonal components in this case.

The tonality is then calculated as

$$K = C\omega_N^{0.79}\sqrt{\sum_{i=0}^{M-1}\left[\omega_1(\Delta z_i)\omega_2(f_i)\omega_3(L_{\Delta,i})\right]^2}^{\,0.29} .$$

where $$\omega_1(\Delta z_i) = \left(\frac{0.13}{\Delta z / Bark + 0.13}\right)^{1/0.29}$$

is a term modeling the dependency of the tonality on the bandwidth Δz of the tonal component.

$$\omega_2(f_i) = \frac{1}{\sqrt{1 + 0.2(f_i / 0.7kHz + 0.7kHz / f_i)^2}}$$

represents the dependency on the frequency f_i of the component. The maximum tonality impression is at 700 Hz.

$$\omega_3(L_{\Delta,i}) = 1 - e^{-\frac{L_{\Delta,i}}{15dB}}$$

calculates the influence of the sound level excess.

The term $$\omega_N = 1 - \frac{N_N}{N}$$

accounts for the ratio of the loudness of the signals *with* its tonal components N and the loudness *without* the tonal components N_N. For the loudness calculation, the HEAD algorithm is used, as it is also based on the publication by W. Aures [1].
The constant C standardizes the result, so that a 1 kHz sine tone at 60 dB has a tonality of 1 tu.
An interesting aspect of tonality is that with an increasing tonality, the euphony of a sound according to Aures increases ([1], page 75), but on the other hand, the degree of annoyance increases, too. The subjective balance of euphony vs. annoyance is very situation-dependent.

---

[1] In order to detect narrow-band noise the level of critical bands with 50% overlap is determined. A critical band is considered as tonal, if its level is at least 7dB higher then the level of the adjacent critical bands. The criterion is adopted from [1], chapter 2.4.2.

Application and Examples

The diagrams in figure 5 show the analysis results for two measurements of sounds of electric generators. The first generator sound (red curve) contains several very distinct tonal components, which are clearly visible in the averaged FFT analysis (left diagram). The tonality analysis of this sound (right diagram) shows an average value of 0.25 tu. The tonal components of the second generator sound (green curve) are considerably less distinct. This is reflected by the tonality analysis in the right diagram: The average value of the analysis is only about 0.05 tu.

Since the annoyance level of noise increases with increasing tonality, the second version of the generator is to be preferred to the first one.

The averaged single values of the tonality analysis in the above example are very suitable, for example, as a measure for a final acoustic test. Using these single values, a classification of acceptable and unacceptable sounds is possible. However, it is important to note that no general threshold value can be defined for this classification. The value that should not be exceeded by a sound to be rated acceptable depends on the type of sound and the application situation.

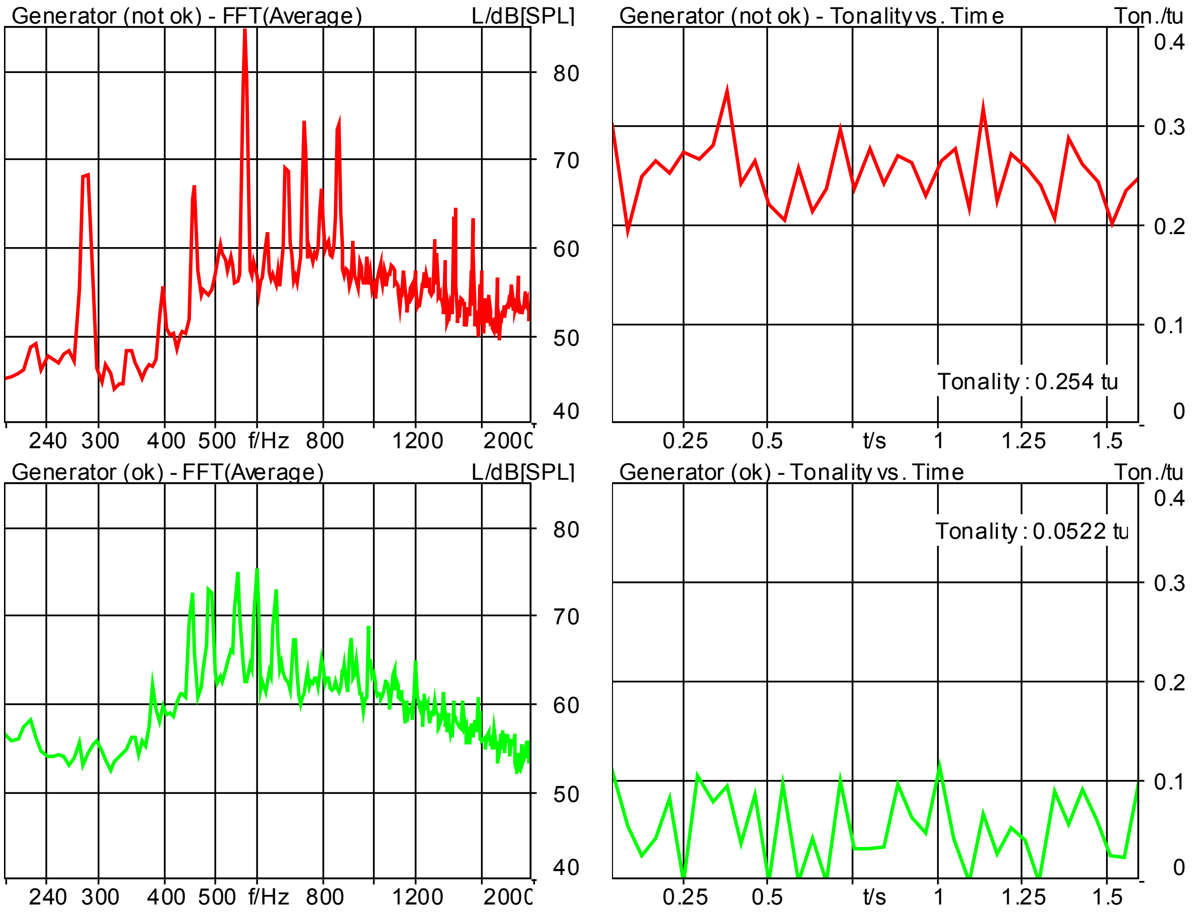

Figure 5: FFT (average) and tonality analysis of two generator sounds

Calculating Tonality according to DIN 45681

The analysis "tonality DIN 45681" serves the determination of the tonal components of noise and determination of 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-dependant masking effect. From the maximum difference, a tonal adjustment value between 0 and 6 dB (in steps of 1 dB) is determined and assigned.

Level Difference	Tonal Adjustment
> 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
- Maximum level difference and frequency of this tone as single value
- Tonal adjustment as single value

In the analyses tonality DIN 45681 vs. time or vs. 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.
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

$$\frac{\text{sampling rate}}{\text{DFT length}} = \frac{44100\,\text{Hz}}{8192} = 5{,}38\,\text{Hz}$$

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.

Notes

For the applications described in this Application Note, you need the ArtemiS basic version (code 4600) and the Psychoacoustics module ATP 02 (code 4602). For the roughness analysis based on the hearing model, you also need the Advanced Psychoacoustics module ATP 06 (code 4606).

Do you have questions for the author? Contact us at imke.hauswirth@head-acoustics.de.
We are looking forward to your feedback!

Literature

[1] W. Aures
Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse (in German)
Dissertation, TU München, 1984

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

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

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

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

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

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

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

[9] E. Terhardt, G. Stoll, M. Seewann
Algorithms for extraction of pitch and pitch salience from complex tonal signals
J. Acoust. Soc. Am. 71 (3) 1982, 679-688